

IMPACT OF FOUR WRF MODIFICATIONS UPON EIGHT NOR'EASTER
SIMULATIONS

by

STEPHEN DAVID NICHOLLS

A dissertation submitted to the
Graduate School-New Brunswick
Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Doctor of Philosophy

Graduate Program in Atmospheric Science

Written under the direction of

Dr. Steven G. Decker

And approved by

New Brunswick, New Jersey

OCTOBER 2012

ABSTRACT OF THE DISSERTATION

Impact of Four WRF Modifications upon Eight Nor'easter Simulations

By STEPHEN DAVID NICHOLLS

Dissertation Director:

Dr. Steven G. Decker

This dissertation investigated the impact of four modifications to the Weather Research and Forecasting Model (WRF) model during eight nor'easter simulations. Specifically these modification include: 1) Different WRF model versions, 2) Usage of different bulk microphysics schemes created between 1983-2011, 3) Assimilation of radio occultation data, and 4) Fully coupling WRF to a dynamic ocean model. Model simulations were conducted for 180 hours, starting roughly 72 hours prior to the first precipitation impacts in the highly populated Mid-Atlantic US and associated cyclogenesis. Simulation accuracy was assessed by comparing each simulation to Global Forecasting System model analysis.

Despite various updates, errors in both storm track and simulated storm intensity were highest in the newest WRF version and were strongly associated with mid-tropospheric heat release. Error analysis of WRF-version simulations revealed the newest WRF model version (WRF 3.3) had worst overall simulation accuracy due to errors in simulated winds, mid-tropospheric latent heat release and similar dynamical fields,

whereas WRF 3.2 was best. Comparison of simulations using different microphysics parameterization revealed both storm tracks and maximum cyclone intensity revealed little to no variation between schemes due to their common programming heritage. Error analysis of the local storm environment revealed simulations little impact from the inclusion of graupel, however the newer microphysics parameterization tended to be more accurate. In contrast, for the entire environment (nor'easter and background) the newest BMPS scheme only performed on-par with the oldest BMPS within the inner most model domains. Improvements to both storm track and overall nor'easter simulation accuracy were typically inversely proportional to the data assimilation period length and was strongly sensitive to cyclone-to-sounding distance and stratospheric data assimilation errors. Simulation accuracy however was not proportional to the total number of assimilated observations. Assimilation of radio occultation data and radiosonde data were found to lead to further decreases in model simulation errors. Finally, coupling WRF to an ocean model produced no notable changes in storm track, slightly improved simulations of cyclone intensity, and marginally better simulations of the local storm environment (54.3% of periods). Impacts from ocean-atmosphere model coupling were limited to below 500 hPa.

Acknowledgements

I express my deepest gratitude and thanks to my research advisor, Dr. Steven G. Decker, for his guidance and support over the last four years which helped make this research possible. My thanks also goes to Drs. Karen Mohr, Dale Haidvogel, and Mark Miller, all of whom reviewed my thesis and provided both their support and vital input. I also wish to thank both the NASA Graduate Student Summer Program (GSSP) for the learning experience afforded by the first author's internship at NASA Goddard Space Flight Center and Dr. Wei-Kuo Tao for his mentorship. In addition during this internship I received much guidance from both Drs. Roger Shi and Steve Lang, whose help and advice were vital to the WRF microphysics chapter of this thesis. I am grateful to both Drs. John Warner and John Wilkin for their invaluable and timely assistance with my research concerning WRF-ROMS coupling and the COAWST model. I express my appreciation to Drs. Ingrid Moore, Bill Kuo, and Ansetty Rao for their constructive and useful feedback concerning the assimilation of COSMIC radio occultation data into WRF. I also acknowledge the help of Dr. Steven Cavallo for his valuable assistance with modifying the RRTM radiation scheme for use in WRF 3.1.1. Finally, I wish to thank my fellow graduate students, professors, and staff at Rutgers University as well as my family for their support and encouragement.

Financial support of this thesis was provided in part by the New Jersey Agricultural Experiment Station and by the 2011 Graduate Student Summer Program (GSSP), the Joint Center for Earth Systems Technology (JCET), and the University of Maryland Baltimore County (UMBC).

Table of contents

ABSTRACT OF THE DISSERTATION	ii
Acknowledgements.....	iv
Table of contents.....	v
List of tables.....	viii
List of illustrations	xi
1. Introduction.....	1
1.1. Introduction to nor'easters	1
1.2. Introduction to WRF model versions	1
1.3. Introduction to microphysics in WRF.....	5
1.4. Introduction to radio occultation	8
1.5. Introduction to ocean-atmosphere coupling and COAWST	11
2. Evaluation of nor'easter simulations from different versions of WRF	15
2.1. Chapter abstract.....	15
2.2. Methods.....	16
2.2.1. Study design	16
2.2.2. Verification and analysis techniques	18
2.3. Results	20
2.3.1. Comparison to Cavallo et al. (2010) local potential temperature tendencies..	20
2.3.2 WRF-version analysis.....	21
2.4. Conclusions	31
2.5. Future work	33
3. Influence of microphysics schemes upon numerical simulations of nor'easters	49

3.1 Chapter abstract.....	49
3.2. Methods.....	50
3.2.1. Study design	50
3.2.2. Verification and analysis techniques	52
3.3. Results	54
3.3.1. Five-class microphysical schemes.....	54
3.3.2. Six-class microphysical schemes.....	63
3.4. Conclusions	72
3.5. Future work	76
4. Impact of cycled assimilation of radio occultation data on nor'easter simulations.....	102
4.1 Chapter abstract.....	102
4.2. Methods.....	103
4.2.1. Study design	103
4.2.2. Verification and analysis techniques	106
4.3. Results	107
4.3.1. COSMIC observations and analysis increments	107
4.3.2. WRF-COSMIC run analysis.....	108
4.3.3. Comparison to larger COSMIC assimilation window and radiosondes.....	117
4.4. Conclusions	123
4.5. Future work	127
5. Impact of coupling an ocean model to WRF nor'easter simulations.....	149
5.1 Chapter abstract.....	149
5.2. Methods.....	150

5.2.1. COAWST Overview.....	150
5.2.2. WRF configuration	150
5.2.3. ROMS configuration	152
5.2.4. Cases for analysis, model run description	153
5.2.5. Verification and analysis techniques	155
5.3. Results	157
5.3.1 Simulated SST comparison	157
5.3.2. WRF-ocean coupling analysis	158
5.4. Conclusions	166
5.5. Future work	168
Bibliography	185

List of tables

Table 1: Nor'easter case list. Column 3 shows the dates which each event impacted the northeastern US, while the last two columns denote the first and last times for each model run.	35
Table 2: GMA-relative storm track bias analysis. Values indicate the total the number six-hour time periods from all cases where each model exhibited the shown track bias relative to GMA at each time period.....	36
Table 3: Performance analysis relative to W311. Values denote the number and percentage of 6-hour periods where the energy norm or RMSE from the indicated simulations was lower than W32. The last section indicates the number of periods where COSMIC-based energy norms were lower than W32 within a 600 km box centered on GMA and each simulated storm, respectively. Bolded values indicate the best WRF simulation (as compared to W311) for each analysis component.	37
Table 4: Microphysics schemes used in the present study and the included microphysics mixing ratios denoted with an "X".	77
Table 5: Nor'easter case list. The NESIS number is included for storm severity reference. The last two columns denote the first and last times for each model run.	78
Table 6: GMA-relative storm track bias analysis for five-class microphysics schemes. Values indicate the total the number six-hour time periods from all cases where each simulation exhibited the shown track bias relative to GMA at each time period.	79
Table 7: Values denote the number and percentage of 6-hour periods where the energy norm or RMSE was lower than WSM5 for each simulation. The last section indicates the number of periods where BMPS-based energy norms were lower than WSM5 within a 600 km box centered on GMA and each simulated storm, respectively. Bolded values indicate the best WRF simulation (as compared to WSM5) for each analysis component.	80
Table 8: Same as Table 6, except for six-class microphysics schemes.	81

Table 9: Values denote the number and percentage of 6-hour periods where the energy norm or RMSE was lower than WSM6 for each simulation. The last section indicates the number of periods where BMPS-based energy norms were lower than WSM6 within a 600 km box centered on GMA and each simulated storm, respectively. Bolded values indicate the best WRF simulation (as compared to WSM6) for each analysis component. 82

Table 10: Nor'easter case list. Column 3 shows the dates which each event impacted the northeastern US, while the last two columns denote the first and last times for each model run. 129

Table 11: GMA-relative storm track bias analysis. Values indicate the total the number six-hour time periods from all cases where each COSMIC model exhibited the shown track bias relative to GMA at each time period. 130

Table 12: Performance analysis relative to W32. Values denote the number and percentage of 6-hour periods where the energy norm or RMSE from the indicated simulations was lower than W32. The last section indicates the number of periods where COSMIC-based energy norms were lower than W32 within a 600 km box centered on GMA and each simulated storm, respectively. Bolded values indicate the best COSMIC simulation for each analysis component. 131

Table 13: Same as Table 11, except it includes only the 180-hr model runs and W32.. 132

Table 14: Same as Table 3, except it includes only full assimilation simulations. Unlike Table 12, this analysis included runs with a ± 1.5 hour COSMIC assimilation window and radiosonde assimilation runs (noted as RAOB). The ± 1 hour window COSMIC assimilation is the same simulation as used in Table 12. Bolded values indicate the best performing data assimilation run for each analysis component..... 133

Table 15: Nor'easter case list. Column 3 shows the dates which each event impacted the northeastern US, while the last two columns denote the first and last times for each model run. Due to missing HYCOM data, cases 2 and 3 are not analyzed. 169

Table 16: GMA-relative storm track bias analysis. Values indicate the total the number six-hour time periods from all cases where each simulation exhibited the shown track bias relative to GMA at each time period. 170

Table 17: Performance analysis relative to W321. Values denote the number and percentage of 6-hour periods where the energy norm or RMSE from the indicated simulations was lower than W321. The last section indicates the number of periods where WRF model run energy norms were lower than W321 within a 600-km box centered on GMA and each simulated storm, respectively. Bolded values indicate the best WRF simulation (as compared to W321) for each analysis component. 171

List of illustrations

Fig. 1: Comparison of model domain and difference in composited, six-hour potential temperature tendency forecasts between WRF and the GFS. Panels a and b depict the model grid configuration from Cavallo et al. (2010; their Fig 1a) and the present study, respectively. Panel c shows various heating rates from Cavallo et al. (2010; their Fig 1b); potential temperature tendency is shown as the right-most, dotted line. Panel d shows the eight-case composite of potential temperature tendency from each WRF model version for domain 1.....	38
Fig. 2: Storm tracks from GMA and each WRF version run. Line legend is shown on the upper-left of each plot. Shown symbols indicate simulated storm position every six hours. White numbers indicate case number.	39
Fig. 3: WRF forecasted storm position bias as compared to GMA for all eight cases. The black arrow in panel 1 indicates the GMA storm motion direction for all panels. Shown symbols represent WRF position bias every six hours (smaller symbols) and their mean (large symbols).....	40
Fig. 4: Various plots from (a,d) GMA, (b,e) W311, and (c,f) W33 (c, f). (a-c) 300-hPa wind speed (fills, knots) and SLP (contours, hPa) and (d-f) 500-hPa positive relative vorticity (fills, 10^{-5} s^{-1}) and 500-hPa geopotential height (contours, m) on 18 UTC 16 March 2007.	41
Fig. 5: Various plots from (a,d) GMA, (b,e,g,i) W311, (h,k) W32 (h, k) and (c,f,i,l) W33. (a-f) 500-hPa positive relative vorticity (fills, 10^{-5} s^{-1}) and 500-hPa geopotential height (fills, m) on 28 February 2009 12UTC (a-c) and 12 UTC 2 March 2009 (d-f). (g-l) 500-hPa latent heating (fills, K day^{-1}) and sea-level pressure (contours, hPa) on 28 February 2009 12UTC (g-i) and 12 UTC 2 March 2009 (j-l).	42
Fig. 6: (a-d) 500-hPa latent heating (fills, K day^{-1}) and sea-level pressure (contours, hPa) and (e-h) 300-hPa winds (fills, knots) and sea-level pressure (contours, hPa). Upper panels (a, b, e, f) are from 15 October 2009 at 12UTC and the lower panels (c, d, g, h) are from 18UTC the same day.	43

Fig. 7: Domain 2 differences in energy norms for all cases within 300 km of the GMA storm center (thin lines) and each model simulated center (thick lines). Shown differences are relative to WRF 3.1.1. Positive values denote improvement over WRF 3.1.1.	44
Fig. 8: Domain 2, time-averaged energy norm for all cases within 300 km of the GFS analysis storm center (thin lines) and each model simulated center (thick lines).	45
Fig. 9: 500-hPa wind speed plotted for (a, c) W311 (grey) and GMA (black) and (b, d) W32 (grey) and GMA (black). (a, b) are on 12 UTC 16 April 2007 and (c,d) are for 12 UTC 5 February 2010.	46
Fig. 10: Energy norm differences within the entirety of domains 1, 2, and 3 for (a, c, e) case 5 and (b, d, e) case 8. Shown differences are relative to WRF 3.1.1. Positive values denote improvement over WRF 3.1.1.	47
Fig. 11: Difference in domain 2 (a) energy norm, (b) sea-level pressure, (c) 850-hPa temperature, (d) 500-geopotential height, (e) 300-hPa winds, (f) 2-PVU potential temperature for case 4. Shown differences are relative to WRF 3.1.1. Positive values denote improvement over WRF 3.1.1.	48
Fig. 12: Nested WRF configuration used in simulations. Horizontal resolution for domains 1, 2, 3, and 4 are 45, 15, 5, and 1.667 km, respectively.	83
Fig. 13: Storm tracks from GMA and the WRF 6-class microphysics runs. Line legend is shown on the upper-left of each plot. Shown symbols indicate simulated storm position every six hours. White numbers indicate case number.	84
Fig. 14: WRF five-class microphysics run storm position bias as compared to GMA for all eight cases. The black arrow in panel 1 indicates the GMA storm motion direction for all panels. Shown symbols represent WRF position bias every six hours (smaller symbols) and their mean (large symbols). Case number is indicated with a white number.	85

Fig. 15: 500-hPa positive relative vorticity (fills; 10^{-5} s^{-1}) and 500-hPa geopotential height (contours; m) on (a-c) 00 UTC 16 March 2007 and (d-f) 06 UTC 2 March 2009. 86

Fig. 16: Domain 4, column-integrated mixing ratios (kg m^{-2}) on 6 Feb 2010 at 06UTC. Shown abbreviations are QC = cloud water mixing ratio, QI = cloud ice mixing ratio, QR = rain mixing ratio, and QS = snow mixing ratio. All shown values are in kg m^{-2} except QI which is scaled by 10^1 87

Fig. 17: Domain 4 time-averaged, case composite mixing ratios (kg kg^{-1}) from all 5-class microphysics scheme model runs. The eight-case average mixing ratio and the highest and lowest mean mixing ratio are indicated by thick and thin lines, respectively. Shown mixing ratio composites include those for (a) water vapor, (b) cloud water, (c) cloud ice, (d) rain and (e) snow. 88

Fig. 18: Domain 2 energy norm differences for all cases within 300 km of the GMA storm center (thin lines) and each model simulated center (thick lines). Shown differences are relative to WSM5 with positive values denoting improvement..... 89

Fig. 19: 180-hr averaged domain 2 energy norm profiles for all cases within 300km of both GMA (thin lines) and each model simulated center (thick lines). The line legend for all 5-Class microphysics schemes is shown in the top right of each panel..... 90

Fig. 20: Energy norm differences within the entirety of domains 1, 2, 3 and 5 for (a, c, e, g) cases 3 and (b, d, e, h) 8. Shown differences are relative to WSM5. Positive values denote improvement over WSM5. 91

Fig. 21: Difference in domain 2 (a) energy norm, (b) sea-level pressure, (c) 850-hPa temperature, (d) 500-geopotential height, (e) 300-hPa winds, (f) 2-PVU potential temperature for case 6. Shown differences are relative to WSM5. Positive values denote improvement over WSM5..... 92

Fig. 22: Sea-level pressure based storm tracks for all eight cases from GMA and each of the six-class BMPSSs. Line legend is shown on the upper-left of each plot. Storm position is indicated once every six hours with the shown dots. 93

Fig. 23: WRF six-class microphysics run storm position bias as compared to GMA for all eight cases. The black arrow in panel 1 indicates the GMA storm motion direction for all panels. Shown symbols represent WRF position bias every six hours (smaller symbols) and their mean (large symbols). Case number is indicated with a white number. 94

Fig. 24: 500-hPa positive relative vorticity (fills; 10^{-5} s^{-1}) and 500-hPa geopotential height (contours; m) on (a-c) 00 UTC 16 March 2007 and (d-f) 06 UTC 2 March 2009. Shown plots are from (a,e) GMA, (b,f), WSM6, (c,g) Lin, and (d,h) NGod. 95

Fig. 25: Domain 4, column-integrated mixing ratios (kg m^{-2}) on 6 Feb 2010 at 06UTC. Shown abbreviations are QC = cloud water mixing ratio, QI = cloud ice mixing ratio, QR = rain mixing ratio, QG = graupel mixing ratio, and QS = snow mixing ratio. All shown values are in kg m^{-2} except QI which is scaled by 10^1 96

Fig. 26: Domain 4 time-averaged, case composite mixing ratios from all 6-class microphysics scheme model runs. The eight-case average mixing ratio and the cases with the highest and lowest mean mixing ratio are indicated by thick and thin lines, respectively. 97

Fig. 27: Domain 2 energy norm differences for all cases within 300 km of the GMA storm center (thin lines) and each model simulated center (thick lines). Shown differences are relative to WSM6 with positive values denoting improvement. 98

Fig. 28: 180-hr averaged domain 2 energy norm profiles for all cases within 300km of both GMA (thin lines) and each model simulated center (thick lines). The line legend for all six-class microphysics schemes is shown in the top right of each panel. 99

Fig. 29: Energy norm differences within the entirety of domains 1, 2, 3 and 5 for (a, c, e, g) cases 3 and (b, d, e, h) 8. Shown differences are relative to WSM6. Positive values denote improvement over WSM6. 100

Fig. 30: Difference in domain 2 (a) energy norm, (b) sea-level pressure, (c) 850-hPa temperature, (d) 500-geopotential height, (e) 300-hPa winds, (f) 2-PVU potential temperature for case 6. Shown differences are relative to WSM6. Positive values denote improvement over WSM6. 101

Fig. 31: Nested grid configuration used in simulations. Horizontal grid spacing for domains 1, 2, and 3 are 45, 15, and 5 km, respectively.	134
Fig. 32: Number of COSMIC data assimilations within a ± 1 hour window every three model hours on domain 1. White number indicate case number.	135
Fig. 33: COSMIC analysis increments from (left) 12 UTC 24 November and (right) 18 UTC 2 March 2009. (top) Differences in 500-hPa geopotential height (fills, m) and sea-level pressure (contours, hPa). (bottom) Differences in 850-hPa water vapor (fills, g kg^{-1}) and temperature (contours, K). Triangles denote COMSIC profile locations.	136
Fig. 34: Storm tracks from GMA and each COMSIC assimilation run. Line legend is shown on the upper-left of each plot. Shown symbols indicate simulated storm position every six hours. White numbers indicate case number.	137
Fig. 35: WRF forecasted storm position bias as compared to GMA for all eight cases. The black arrow in panel 1 indicates the GMA storm motion direction for all panels. Shown symbols represent WRF position bias every six hours (smaller symbols) and their mean (large symbols).	138
Fig. 36: Various plots from (a,d,g) GMA, (e,h) W32, (b) WC-48 hr (B), and (c,f,i) WC-180hr. (a-c) 500-hPa relative vorticity (shaded, s^{-1}) and geopotential height (contours, m) on 12 UTC 16 March 2007. (d-f) 500-hPa geopotential height (shaded, dam), and sea-level pressure (hPa, contours) on 18UTC 28 Feb. 2009. (g-i) same as (d-f) except for 18 UTC 2 March 2009 at 18UTC.	139
Fig. 37: (a-d) 500-hPa latent heating (fills, K day^{-1}) and sea-level pressure (contours, hPa) and (e-h) 300-hPa winds (fills, knots) and sea-level pressure (contours, hPa) (E-H). (a,b,e,f) are from 12 UTC 15 October 2009 and (c,d,g,h) are from 18 UTC the same day.	140
Fig. 38: Domain 2 differences in COSMIC run energy norms for all cases within 300 km of the GFS analysis storm center (thin lines) and each model simulated center (thick lines). Shown differences are relative to WRF 3.2. Positive values denote improvement over WRF 3.2.	141

Fig. 39: Domain 2, time-averaged energy norm for all cases with 300 km of the GFS analysis storm center (thin lines) and each model simulated center (thick lines).....	142
Fig. 40: Energy norm differences in COSMIC runs within the entirety of domains 1, 2, and 3 for (a, c, e) case 1 and (b, d, e) case 4. Shown differences are relative to W32. Positive values denote improvement over W32.....	143
Fig. 41: Difference in domain 2 (a) energy norm, (b) sea-level pressure, (c) 850-hPa temperature, (d) 500-geopotential height, (e) 300-hPa winds, and (f) 2PVU potential temperature for case 3. Shown differences are relative to W32. Positive values denote improvement over W32.	144
Fig. 42: WRF forecasted storm position bias as compared to GMA for all eight cases. The black arrow in panel 1 indicates the GMA storm motion direction for all panels. Shown symbols represent WRF position bias every six hours (smaller symbols) and their mean (large symbols).....	145
Fig. 43: Domain 2 differences in 180 hour COSMIC (1 and 1.5 hour window) and radiosonde assimilation energy norms for all cases with 300 km of the GFS analysis storm center (thin lines) and each model simulated center (thick lines). Shown differences are relative to WRF 3.2. Positive values denote improvement over WRF 3.2.	146
Fig. 44: Energy norm differences in 180 hour COSMIC (1 and 1.5 hour window) and radiosonde assimilation energy norms COSMIC runs within domains 1, 2, and 3 for case 1 (a,c,e) and case 4 (b,d,e). Shown differences are relative to W32. Positive values denote improvement over W32.	147
Fig. 45: Difference in domain 2, 180 hour COSMIC (1 and 1.5 hour window) and radiosonde assimilation (a) energy norm, (b) sea-level pressure, (c) 850-hPa temperature, (d) 500-geopotential height, (e) 300-hPa wind, and (f) 2PVU potential temperature for case 3. Shown differences are relative to W32. Positive values denote improvement over W32.....	148

Fig. 46: All available COAWST configurations involving the WRF model and exchanged data fields. (a) Is WRF only, (b) WRF-ROMS coupling, (C) WRF-ROMS-SWAN coupling, (d) WRF-ROMS-SWAN-CSTMS coupling. Adapted from Fig. 5 of Warner et al. (2010). 172

Fig. 47: (a) WRF model configuration where domains 1, 2, and 3 have 45, 15, and 5 km grid spacing, respectively. (b) ROMS model configuration superimposed on WRF model domain 2 with 8 km average grid spacing. 173

Fig. 48: Simulated sea surface temperatures (K) from W321, WPol, and COAWST at (a-c) 18UTC 2 February 2010, (d-f) 06UTC 10 February 2010 and (g-i) their difference. 174

Fig. 49: Storm tracks from GMA and WRF runs. Line legend is shown on the upper-left of each plot. Shown symbols indicate simulated storm position every six hours. White numbers indicate case number. 175

Fig. 50: WRF forecasted storm position bias as compared to GMA for all eight cases. The black arrow in panel 1 indicates the GMA storm motion direction for all panels. Shown symbols represent WRF position bias every six hours (smaller symbols) and their mean (large symbols). Case number is indicated with a white number. 176

Fig. 51: Various plots from GMA, W321 and COAWST. (a-f) 850-hPa equivalent potential temperature (fills, K) and sea-level pressure (contours) and (g-l) 500-hPa positive relative vorticity (fills, 10^{-5} s^{-1}) and 500-hPa geopotential height (fills, m). (a-c, g-i) are on 28 February 2009 12UTC and (d-f, j-l) are on 12 UTC 2 March 2009. 177

Fig. 52: Various plots from GMA, W321 and COAWST. (a-c) 300-hPa wind speed (knots, fills) and sea-level pressure (hPa, contours), (d-f) 500-hPa positive relative vorticity (fills, 10^{-5} s^{-1}) and 500-hPa geopotential height (fills, m), and (g-i) 850-hPa equivalent potential temperature (K, fills) and sea-level pressure (hPa, contours). All plots are for 00UTC 14 March 2010. 178

Fig. 53: (a-b) 850-hPa equivalent potential temperature (fills, K) and sea-level pressure (hPa, contours), (c-d) 500-hPa positive relative vorticity (fills, 10^{-5} s^{-1} and 500-hPa geopotential height (m, contours), and (e-f) 500-hPa latent heating (fills, K day^{-1}) and sea-level pressure (contours, hPa). All plots are from 15 October 2009 at 12UTC. 179

Fig. 54: Domain 2 energy norm differences for all cases within 300 km of the GMA storm center (thin lines) and each model simulated center (thick lines). Shown differences are relative to W321 with positive values denoting improvement. 180

Fig. 55: Domain 2, time-averaged energy norm for cases 1, 4, 6 and 7 with 300 km of the GFS analysis storm center (thin lines) and each model simulated center (thick lines). . 181

Fig. 56: Energy norm differences within the entirety of domains 1, 2, and 3 for (a, c, e) case 4 and (b, d, e) case 5. Shown differences are relative to W321. Positive values denote improvement over W321..... 182

Fig. 57: SST (fills, K) and SLP (contours, hPa) on 00 UTC 17 October 2009 from (a) GMA, (b) W321, (c) WPol, and (d) COAWST..... 183

Fig. 58: Difference in domain 2 (a) energy norm, (b) sea-level pressure, (c) 850-hPa temperature, (d) 500-geopotential height, (e) 300-hPa winds, (f) 2-PVU potential temperature for case 1. Shown differences are relative to W321. Positive values denote improvement over W321. 184

1. Introduction

1.1. Introduction to nor'easters

To those living in the Northeast United States (NEUS), frozen precipitation is not an uncommon occurrence in association with extratropical cyclones, especially during the winter months. The most infamous cyclone type is the so called “nor’easter”. We define a “nor’easter” as a large (~2000 km), wintertime cyclone occurring between October and April, bringing punishing winds, copious precipitation, and the potential for coastal flooding to the Northeastern United States (Kocin and Uccellini 2004; Jacobs et al. 2005; Ashton et al. 2008). Nor’easter-related financial damages can reach several billion US dollars per event (NCDC 2008). Trenberth et al. (2007) noted that during the 20th Century, mid-latitude cyclones (including nor’easters) tracked increasingly poleward, were more infrequent, yet more intense. These changes were attributed to warmer global-average air and sea-surface temperatures and increased atmospheric water vapor content.

Given climate model predictions of a 3°C warming during the 21st Century, it is likely that future nor’easters will be more intense, yet less frequent than at present. Should these predictions prove valid, the potential disruption and damage incurred per future nor’easter event will only increase. To mitigate this risk, we need to improve upon current numerical weather prediction models (NWPMs) to improve their accuracy and coherence at longer time scales.

1.2. Introduction to WRF model versions

Since 1950, when the first numerical weather forecast was completed on the Electronic Numerical Integrator and Computer server, NWPMs have become integral to

atmospheric science. As computer technology has advanced, so have the complexity and capability of NWPMs. Present-day NWPMs, such as the Weather Research and Forecasting (WRF) model, integrate complex, highly-detailed, non-hydrostatic equations to produce simulations with levels of accuracy and resolution that were inconceivable in 1950 (Michalakes et al. 2004; Skamarock et al. 2008; Wang et al. 2012). More recent NWPM advances have seen the greatest improvements over regions of high terrain (Mass et al. 2002) in the positioning and intensity of convection (Weisman et al. 2008). Despite over 62 years of model advances, NWPM simulations remain imperfect. Due to these imperfections, NWPMs will continue to be updated and refined as new errors are detected, capabilities added, or parameterizations implemented. While the modification of individual NWPMs (Cavallo et al. 2010; Shi et al. 2010) and inter-forecast office model comparisons (i.e., Buizza et al. 2005) have been well-documented, few studies address how individual model performance varies with version release. To investigate this less explored topic, we focus on how WRF model simulations of nor'easters varied amongst WRF versions 3.1.1, 3.2, and 3.3 in chapter 2.

The WRF model system itself is comprised of three main components: the WRF preprocessing system (WPS), WRF data assimilation, and the Advanced Research WRF (ARW) solver (NCAR 2012e). First, the WPS ingests gridded model data from an external source (e.g., Global Forecasting System [GFS]) and converts it into WRF legible format. Second, the WRF data assimilation program ingests additional observations, converts them to WRF legible format, and then perturbs the WRF input data while accounting for observation error. Finally, once all data is processed, the ARW solver carries out the forward integration.

The evolution of WRF from version 3.1.1 to 3.3 occurred over a roughly 20 month period. During this time, model coders addressed several minor tweaks (e.g., correcting units for snow height), expanded functionality (e.g., additional model assimilation tables) and applied a major correction to a well-used longwave radiation parameterization (NCAR 2012a, b, c, d).

Between versions 3.1.1 and 3.2, WRF coders made five notable modifications. Version 3.2 introduced an independent, land-surface coupled, building energy model parameterization (Chen et al. 2011). Based upon user need and computation expense, WRF users either parameterized urban surface processes or ran a sub-grid scale, multi-layer urban canopy model to describe indoor-outdoor heat exchanges and their interaction with the atmospheric boundary layer (Chen et al. 2011). Also introduced was full-coupling of WRF to the Pollard et al. (1972) single-column, mixed-layer ocean model. Until this release, sea surface temperatures (SSTs) were either fixed or updated from an external source. The new coupled model, fitted standard sub-surface mean temperature profiles to SST data (Pollard et al. 1972) and then adjusted ocean mixed-layer height in response to imposed wind stress. The ocean model however neglects horizontal advection (Pollard et al. 1972, NCAR 2012e). Next, an error with the WRF single- and double-moment microphysics schemes (Hong et al. 2004; Lim and Hong 2010) for simulations with time steps longer than two minutes was corrected. Prior to this fix, these schemes passed erroneous precipitation data to the land surface models (NCAR 2012c). Another addition in version 3.2 was the Milbrandt and Yau, seven-class, double-moment microphysics scheme (Milbrandt and Yau 2005). Unlike other double-moment schemes, it has distinct bins for hail and graupel (Milbrandt and Yau 2005; NCAR 2012a). Last, a

significant correction was made to the Rapid Radiative Transfer Model (RRTM) longwave radiation scheme. Cavallo et al. (2010) discovered a -15 K day^{-1} longwave heating rate cold bias above 100 hPa and slight warming near the tropopause (see Fig. 1b), which limited model stability. They reduced these biases to within $\pm 0.5 \text{ K day}^{-1}$ of the standard cooling rates by adding additional model layers (4hPa interval and smaller), fitting standard, mean temperature profiles (adjusted for latitude) and fixing a minimum water vapor content of 5 ppmv above the model top.

Model coders made six notable changes between WRF versions 3.2 and 3.3. Version 3.3 introduced a stochastic kinetic-energy backscatter scheme for large-scale eddy simulations near the surface (Shutts 2005, NCAR 2012 a, e). This scheme injects sub-grid scale energy back into the model to offset unrealistically large advection and horizontal energy diffusion errors in NWPMs (Shutts 2005). Once injected, energy transfers can reach explicitly resolved scales and resulted in both increased forecast spread amongst ECMWF ensemble model members and forecast skill (Shutts 2005). Next, model developers included the Stony-Brook University Lin Scheme which is a five-class, single-moment microphysics scheme (Lin and Colle 2011). Unlike similar parameterizations, it explicitly diagnoses riming intensity and produces a continuous mixed-phase precipitation (precipitating ice) spectrum ranging from ice particles through graupel and snow. Additionally, the relationship between temperature and riming to the area, mass, and fall velocity of precipitating ice is explicitly calculated. Version 3.3 also contained a second WRF shallow convection parameterization derived from the Community Earth System climate model (Park and Bretherton 2009). This parameterization updated model treatment of lateral mixing about cumulus updrafts and

cloud to sub-cloud interactions based upon single-column shallow cumulus simulations of Brown et al. (2002). Another new release with WRF 3.3 was the Goddard longwave scheme (Shi et al. 2010). This scheme when combined with the existing Goddard shortwave model parameterization (Chou and Suarez 1999) afforded simulations of the aerosol indirect effect. Finally, the RRTM Global Climate Model parameterization was updated in lieu of Cavallo et al. (2010) to correct an identical cold bias as exhibited by the RRTM.

1.3. Introduction to microphysics in WRF

Prognostic models and their associated microphysics parameterizations have become increasingly complex, accurate, and computationally expensive. Modern numerical weather prediction models (NWPMs), such as the Weather Research and Forecasting (WRF) model (Michalakes et al. 2004; Skamarock et al. 2008; Wang et al. 2012), offer a numerous microphysics configurations ranging from the simple, warm-rain only Kessler scheme (Kessler 1969) to the double-moment, six-class Morrison scheme (Morrison et al. 2009). Cloud microphysics has shown to be important to NWPM accuracy and have been the focus of at least 36 major studies as noted by Tao et al. (2011). All but one study mentioned in Tao et al. (2011) involved idealized simulations, hurricanes, or mid-latitude convection. For chapter 3, we will focus on an unfamiliar topic: Wintertime, mid-latitude cyclones and specifically nor'easters.

The effect of BMPSs on NWPM forecasts of nor'easters or mid-latitude cyclones in general is not well-investigated in the literature. For instance, Reisner et al. (1998) evaluated Mesoscale Model Version 5 (MM5) simulations of super cooled water in

winter storms from the Winter Icing and Storms Project (WISP). Between model runs, only the BMPSs varied. Their results found MM5 simulations utilizing double-moment BMPSs more closely matched WISP observations than single-moment schemes. However, single-moment BMPS simulation performance was noted to be further enhanced when a diagnostic equation was used to calculate the snow-size distribution intercept. To obtain realistic simulations of super cooled liquid water in winter storms, accurate simulations of its kinematic and dynamical structure were required.

More recently Wu and Petty (2010) (hereafter WP10) evaluated WRF simulations using five, mixed phase, six-class BMPSs during four polar-low events (two over Japan, two over the Nordic Sea) using three single-moment and two double-moment BMPSs. Their simulations exhibited near identical storm tracks, but had notable errors in both cloud top temperatures and precipitation fields. The single-moment Lin (Lin et al. 1983) and double-moment Thompson (Thompson et al. 2008) schemes did not produce realistic cloud coverage due to insufficient snow and graupel generation. The double-moment Morrison scheme produced unrealistically high cloud-water concentrations in the lower atmosphere. Overall, WP10 found the WRF single-moment, six-class (WSM6; Hong and Lim 2006) scheme produced marginally superior simulations of cloud and precipitation processes as compared to other BMPSs.

Shi et al. (2010) evaluated WRF model simulations using five and six-class, single-moment BMPSs during both a lake-effect snow event and a 20-22 January 2007 synoptic event. By comparing simulated radar reflectivity and cloud top temperatures to observations they found WRF accurately simulated event onset and termination, cloud coverage, and the size and scope of the lake-effect snow band. In contrast, WRF was

unable to predict the correct snowfall rate due to difficulties in predicting point observations within a mesoscale grid. Finally, five- and six-class BMPS simulations were nearly identical because both cold temperatures and weak vertical velocities were unfavorable for graupel generation.

Similar to WP10, Tao et al. (2011) evaluated WRF model performance using four, six-class BMPSs to simulate Hurricane Katrina. Unlike WP10 only BMPS was run using three different configurations: 2-ice (snow, cloud ice). 3-ice-graupel (snow, cloud ice, graupel), 3-ice-hail (snow, cloud ice, hail). Results indicated that BMPS selection had minimal impact upon storm track, but upwards of a 50-hPa variation in lowest attained sea-level pressure (SLP).

Amongst these previous studies, the present study is unique because nor'easters are meteorological phenomenon of both larger size and multi-scale. The aim of this study is understand whether various aspects of simulated nor'easters (e.g., SLP, storm track, winds, etc.) are sensitive to microphysics schemes and whether this sensitivity is detectable. Furthermore, how various microphysical properties (i.e., mixing ratios) are altered between schemes will also be investigated.

Given predictions of more powerful mid-latitude cyclones over the 21st Century understanding how factors including cloud microphysics potentially influence their genesis, associated precipitation patterns, and latent heating patterns merits investigation. For the current work, we utilized the WRF Advanced Research WRF (WRF-ARW) Model Version 3.3 (Hereafter W33) to evaluate the potential impact of five and six-class bulk microphysics schemes (BMPSs) on WRF simulations of nor'easters.

1.4. Introduction to radio occultation

Presently, the Continental United States relies upon a network of 70 radiosonde stations for mid and upper troposphere initialization in operational weather models (i.e., the Weather Research and Forecasting [WRF] model) (Michalakes et al. 2004; Skamarock et al. 2008; Wang et al. 2012). These data have low spatial (~ 300 km), and temporal resolution (~ 12 hours), potentially limiting their benefits to model initialization during either rapid cyclogenesis or development in data sparse regions. To bypass this problem, forecast centers can assimilate ex-situ microwave profiler data (e.g., the Microwave Sounding Unit; Velden et al. 1991), which, although useful, are unable to fully penetrate cloudy, rain-laden environments typical of mid-latitude cyclones (such as nor'easters). Due to these limitations, neither data source is able to provide usable data from within the local-storm environment, which is vital to nor'easters (Anthes et al. 1983; Kuo et al. 1991, Ren et al. 2004; Eckhardt and Stohl, 2004; Knippertz and Wernli 2009). Global Positioning System radio occultation (GPSRO) attempts to solve these limitations with its global coverage and all-weather capability (Ware et al. 1996; Anthes et al. 2008; Cucurull et al. 2008; Liu et al. 2012). Current GPSRO research has primarily focused on tropical cyclones (Biondi et al. 2011), ionospheric research (Aragon-Angel et al. 2011), monsoons (Huang et al. 2007), climate (Anthes 2011) or reflected GPS signals (Boniface et al. 2011). We will address a less familiar topic: GPSRO assimilation impacts on nor'easter simulations specifically in chapter 4.

Radio occultation involves the derivation of occulted GPS signal refractivity from signal propagation time and Doppler frequency shift measurements relative to an orbiting receiver. Data post-processing uses an Abel transform (Ware et al., 1996; Anthes et al.

2008) and open-loop tracking (Sokolovskiy 2001) to construct vertical profiles of bending angle and refractivity at a tangent point. Profile penetration depends upon the inversion algorithm, receiver signal algorithms, terrain, vertical profile structure, and azimuth angle of the occulted GPS satellite relative to the receiver (Jensen et al. 2003; Anthes et al. 2008). Using these refractivity data, temperature, moisture, and pressure profiles are numerically derived by a numerical weather prediction model via the mathematical relationship

$$N = 77.6 \frac{p}{T} + 3.73 \times 10^5 \frac{e}{T^2} - 4.03 \times 10^7 \frac{n_c}{f^2} \quad (1)$$

where f is frequency (Hz), T is temperature (K), e is vapor pressure (hPa), p is dry pressure (hPa), n_c is electron density (m^{-3}) and N is refractivity.

Active GPSRO missions include the following: Challenging Minisatellite Payload (CHAMP), Communications/Navigation Outage Forecasting System (C/NOFS), Satélite de Aplicaciones Científicas-C (SAC-C), and the Constellation Observing System for Meteorology, Ionosphere, and Climate (COSMIC) (Huang et al. 2007; Ma et al. 2011). Due to their high accuracy ($<0.5\text{K}$ of validation observations), long-term dataset stability ($0.10 \text{ K decade}^{-1}$), and high precision (2x's more precise than radiosondes) these data have been used extensively for both climate monitoring and initializing numerical weather prediction model initialization (Kuo et al. 2005; Poli et al. 2010; Anthes 2011; Lackner et al. 2011; Steiner et al. 2011). To mitigate any possible inter-instrument variability, we focused entirely on the largest GPSRO mission: COSMIC. This joint mission of the National Space Organization of Taiwan and UCAR is comprised of six identical polar-orbiting microsatellites and currently averages 1,800 daily soundings, down from 2,500 at its peak (Cheng et al. 2006; Anthes et al. 2008; Anthes 2011). Data

from COSMIC has high vertical resolution (60-100m) and near zero inter-satellite variance (Poli et al. 2010), yet low horizontal resolution (approx. 200km) (Gorbunov et al. 2004; Anthes et al. 2008; Anthes 2011).

Current COSMIC research has demonstrated its benefits to a wide-range of environment observation and modeling studies. Biondi et al. (2011) used GPSRO bending angle anomalies for convective tower detection within tropical cyclones. Huang et al. (2007) demonstrated that upstream COSMIC data assimilation could improve Indian Monsoon simulations in WRF. Ma et al. (2011) used COSMIC to improve NCEP Regional GSI System short- and medium-range atmospheric river simulations. Healy and Thépaut (2006) attributed an 11% reduction through 10 days in ECMWF stratospheric temperature RMS error to COSMIC. Global Forecasting System (GFS) COSMIC assimilation significantly reduced 500-hPa anomaly correlation scores out to seven days (Cucurull et al. 2008). Anthes et al. (2008) and Liu et al. (2012) attributed lower-troposphere moistening due to COSMIC to significant improvements to five-day WRF simulations of Hurricane Ernesto. These improvements vanish once COSMIC was not assimilated below 6 km (Liu et al. [2012]).

Since its inception, GPSRO has been proven a reliable, accurate, and versatile tool for studying the atmosphere and improving model simulations. Thus far GPSRO-based studies have primarily focused upon tropical cyclones, global model simulations, and the study of the ionosphere. Less focus has been applied to other phenomenon, especially outside of the tropics. In this study, we will explore less familiar territory by assessing how COSMIC data impacted WRF nor'easter simulations.

1.5. Introduction to ocean-atmosphere coupling and COAWST

As computational resources continue to improve, so too will the complexity and detail provided by numerical weather prediction models. Modern atmospheric prognostic models, such as the Weather Research and Forecasting (WRF) Model (Michalakes et al. 2004; Skamarock et al. 2008; Wang et al. 2012), derive their forecasts by integrating complex, highly-detailed, non-hydrostatic equations. Despite its complexity and land-surface coupling (via models such as the NOAH land surface model [Chen and Dudhia 2001]), it did not contain any ocean-atmosphere coupling until April 2009 (WRF version 3.2). Numerous studies have shown that accurate ocean-atmosphere flux and momentum exchanges are vital to simulations of tropical cyclones (Sutyrin and Khain 1984; Bender et al. 1993; Emanuel 1999; Warner et al. 2010; Olabarrieta et al. 2012) and mid-latitude, baroclinically-driven, coastal cyclones (Anthes et al. 1983; Kuo et al. 1991, Ren et al. 2004, Eckhardt and Stohl, 2004; Knippertz and Wernli 2009). Specifically, these fluxes and exchanges were noted to directly key cyclone properties including its propagation, intensity, and maintenance.

To more accurately resolve ocean-atmosphere exchanges during such intense cyclones, several studies have developed or implemented coupled ocean-atmosphere models. An early study investigating ocean-atmosphere dynamics during tropical cyclones was Bender et al. (1993). This study focused on wind-stress-induced sea surface temperature (SST) cooling associated with Hurricane Gloria via coupling of the NOAA Geophysical Fluid Dynamics Laboratory tropical cyclone prediction model to a multilayer primitive equation ocean model. Simulations only varied by storm propagation speed which produced SST cooling of 5.3°C, 3.5°C, and 1.8°C for slow, medium, and

fast propagation, respectively. Reductions in SSTs significantly reduced the total heat flux and in turn both tropical cyclone intensity (7.3 hPa) and maximum wind (2.7 m s^{-1}). Storm track however, was not significantly altered.

More recently, Ren et al. (2004) coupled the Canadian Mesoscale Compressibility Community (CMC2) atmospheric model to the full-physics, 3D Princeton Oceanography Model (POM) to investigate ocean-atmosphere dynamics during the extratropical transition of Hurricane Earl in 1998 and an intense January 2000 winter cyclone dubbed “Superbomb”. These two systems were selected due to comparable size and intensities. Despite their similar characteristics, Hurricane Earl produced greater SST reductions than Superbomb (5 and 1°C , respectively) due to the thicker mixed layer depth in the latter. Similar to Bender et al. (1993), simulations of Earl exhibited no notable storm track changes, however reductions to both storm intensity and maximum wind speed were more muted (3 hPa and 3 m s^{-1} , respectively). Despite its thicker ocean mixed layer, Superbomb still produced similar intensity and wind speed reductions as Earl (4 hPa and 4 m s^{-1}) due to its slower propagation speed.

Using the same coupled model configuration as Ren et al. (2004), Yao et al. (2008) investigated 42 North Atlantic October storms. Their results showed stronger SST cooling ($\sim 6^\circ\text{C}$), slightly stronger cyclone weakening (4-5 hPa), but a similar degree of weakening of 10 m winds ($2\text{-}4 \text{ m s}^{-1}$) as noted Ren et al. (2004). Higher degrees of both SST cooling and increased sea-level pressure (SLP), as compared to Ren et al. (2004), can be attributed to the shallower mixed layer depths that occur in October as opposed to January.

As of WRF version 3.3, three options exist for addressing SSTs: Static SSTs, SST update, or 1D ocean mixed-layer coupling. As WRF requires SST data to function, a static option

requires input from an external source (e.g., Global Forecasting System [GFS] output, 0.5° real-time global sea surface temperature [RTG-SST; Gemmill et al., 2007]), which is held constant for the entire model run. The second option, SST update, requires SST data (such as RTG-SST) for each model boundary condition, and from these data SSTs are updated at user-prescribed intervals and are constant otherwise. Last, the 1D ocean model option couples WRF to the Pollard et al. (1972) 1D ocean mixed layer model. Ocean initialization involves affixing standard ocean temperature profiles to surface SSTs. Initial mixed layer depth is everywhere uniform and user specified. Running this model applies wind stress to the ocean surface creating turbulent mixing throughout the mixed layer. Because this ocean model is one-dimensional, neither advection nor ocean processes such Ekman Transport are simulated. To maintain ocean accuracy, SSTs are updated via the lower boundary condition at user-prescribed intervals, but unlike SST update, SSTs are allowed to change between boundary condition updates in response to ocean mixing. Because these options do not include a fully three-dimensional ocean model or higher resolution SST data, WRF is unable to fully resolve fine-scale SST gradients, leading to heat flux and surface wind stress errors (Chelton et al. 2001; O'Neill et al. 2003; Tikinaga et al. 2005; LaCasse et al. 2008). Such errors have served as motivation for the development of WRF-ocean coupled systems, such as the Coupled Ocean-Atmosphere-Wave-Sediment Transport (COAWST) modeling system (Warner et al. 2010).

The COAWST system couples WRF to the Regional Ocean Modeling System (ROMS; Shchepetkin and McWilliams 2005; Haidvogel et al. 2008), the Simulating Waves Nearshore (SWAN; Booij et al. 1999) wave and the Community Sediment Transport Modeling System (CSTMS; Warner et al. 2008) models via the Model Coupling Toolkit (MCT; Larson et al. 2004). As described by Warner et al. (2010), the COAWST model was specifically designed “to better identify the significant processes affecting our coastlines and how those processes create coastal change.” Since its

development, COAWST-related studies have focused primarily on two main themes: Tropical cyclones (Warner et al. 2010; Olabarrieta et al. 2012), and coastal processes including rip current and surf zone simulations (Kumar et al. 2011; Olabarrieta et al. 2011). Applications of COAWST to baroclinically driven midlatitude cyclones (such as nor'easters) have yet to be addressed.

Given predictions of more powerful mid-latitude cyclones over the 21st Century, improving simulations of the dynamic air-sea interface that underpins nor'easter development will likely prove vital. To investigate the importance of this interface and simulate its overall impact, this study will use COAWST to couple WRF to ROMS during eight nor'easter events and will discuss it in chapter 5.

2. Evaluation of nor'easter simulations from different versions of WRF

2.1. Chapter abstract

Nor'easter simulation accuracy from three Weather Research and Forecasting (WRF) Model versions (3.1.1, 3.2 and 3.3) and WRF 3.1.1 using a Rapid Radiation Transfer Model (RRTM) update were compared. Simulations had identical configurations, ran for 180 hours, and were initialized starting roughly 72 hours prior to the first precipitation impacts in the highly populated Mid-Atlantic US and associated cyclogenesis. A majority of WRF simulations (20 out of 32) exhibited a leftward track bias and lagged Global Forecasting System Model Analysis. Track errors associated with WRF version 3.3 however, were larger (average ~80km) than other WRF runs. Simulated cyclone intensity was strongly associated with mid-tropospheric latent heat release and only 7 out of 32 simulations had minimum sea-level pressure values within 5 hPa of ground truth. Error analysis for both the local-storm environment (within a 600 km nor'easter-centered square) and overall indicated that although simulations from WRF 3.1.1 with the RRTM update and WRF 3.2 showed improvement (up to 74% of the time) the associated increases in model accuracy were modest. Simulation error associated with version 3.3 was at least four times greater than in other WRF simulations. Specifically WRF 3.3 (at worst) exhibited lower simulation error than WRF 3.1.1 and WRF 3.2 only 35% and 31% of the time, respectively. This poor showing by WRF 3.3 was partly associated with increased error in simulated winds, mid-tropospheric latent heat release and similar dynamical fields.

2.2. Methods

2.2.1. Study design

We utilized the WRF Advanced Research WRF which solves fully-compressible, non-hydrostatic, Eulerian equations in terrain-following coordinates (Skamarock et al. 2008; Wang et al. 2012). A nor'easter is a multi-scale phenomenon, which required model grids of sufficient size and resolution to fully capture its dynamics. The resulting model grid configuration (Fig. 1c) includes three domains with two-way interaction, 45, 15, and 5 km grid spacing, respectively, 27 vertical levels, and a 50-hPa model top. This configuration afforded simulation of key pre-cursor synoptic and meso- α -scale phenomena (e.g., jet streaks, short- and long-wave troughs) on outer domains and smaller-scale phenomena (e.g., orographic forcing, latent heating) on inner domains. Boundary conditions were derived from GFS model forecasts ($1^\circ \times 1^\circ$ resolution).

Model simulations were conducted for 180 hours, starting roughly 72 hours prior to the first precipitation impacts in the highly populated Mid-Atlantic US and associated cyclogenesis. This time frame focused attention exclusively on cyclone initiation and its later impact in this region. A 72-hour lead time allowed simulations to spin-up, establish baroclinicity between the cooler eastern United States and warmer Gulf Stream, and simulate latent heating along the expansive (>1000 km) northern edge of the Gulf Stream. All of the above are vital for accurate nor'easter simulations (Kuo et al. 1991; Mote et al. 1997; Yao et al. 2008). Precipitation data from the New Jersey Weather and Climate Network (Robinson 2005) served as a proxy for establishing when each nor'easter first impacted the Mid-Atlantic US. Using these data, model initialization was set as 72 hours prior to the first nor'easter-related 0.5 mm (~ 0.02 inch) precipitation reading. A New

Jersey-centric approach was chosen because of its high population density ($461.6/\text{km}^2$), significant contribution (\$473 billion) to the US gross domestic product, and its relatively central location in the region of interest (United States Census Bureau 2012).

Model parameterizations were selected following a qualitative comparison of many tens of 12-hour, single-domain WRF simulations during a November 2006 nor'easter to GFS model analysis (GMA). The selected parameterizations (see below) were from the model run providing both the best comparison to GMA and the ability to complete a three-domain (see Fig. 1), 180-hour model simulation in less than 12 hours of computational time.

- Longwave radiation: RRTM (Mlawer et al 1997)
- Shortwave radiation: Dudhia (Dudhia 1989)
- Microphysics: Goddard 3-ice, graupel (Lang et al. 2007)
- Surface layer: MM5 similarity (Zhang and Anthes 1982)
- Land surface: NOAH (Chen and Dudhia 2001)
- Boundary layer: BouLac (Bougeault and Lacarrère, 1989)
- Cumulus parameterization: Grell-Devenyi ensemble scheme (Grell and Devenyi 2002)

This study focused on eight nor'easter cases (Table 1) selected based upon their timing, intensity, and track. All cases occurred after an April 2006 satellite launch to correlate with a parallel nor'easter study (Nicholls and Decker 2012) and during the “nor'easter season” (October to April) as defined by Jacobs et al. (2005). To make this study more generalized, we did not select only the most severe nor'easters, but instead aimed for a relatively small but diverse sample. Case severity was measured using the

Northeast Snowfall Impact Scale (NESIS; Kocin and Uccellini 2004). Under NESIS, nor'easters were classified on a scale of 1 (notable) to 5 (extreme) based upon the population impacted, area affected, and event severity. Finally, all cases were required to have precipitated in New Jersey.

To investigate the impact of WRF version, four model runs were completed per case. These runs included simulations from WRF versions 3.1.1, 3.2, 3.3, and version 3.1.1 utilizing the RRTM scheme distributed with version 3.2. The latter simulation focused solely upon the significance of the Cavallo et al. (2010) RRTM update. With the exception of model version, all model runs were configured and compiled identically. For brevity, we will refer to these runs as W311, W32, W33, and W311-MRad, respectively.

2.2.2. Verification and analysis techniques

Validation data was derived from GMA. This data source was chosen because all model domains included data sparse regions where in-situ observations were frequently unavailable, and GMA was easily interpolated to the WRF model grid. Furthermore, the lateral boundary conditions were GFS-based.

Model run analysis was comprised of several parts. For comparison to Cavallo et al. (2010), composited six-hour potential temperature tendency was compared between each WRF version and GFS model forecast. Given the model domain (Fig. 1a) of Cavallo et al. (2010), only results from domain 1 are shown. Storm tracks were determined using local-minima in sea-level pressure (SLP) via an objective, self-developed, WRF-track algorithm similar to that used at the Climate Prediction Center (Serreze 1995; Serreze et al. 1997). To elucidate storm track differences and potential biases each simulation was

compared to GMA every six hours. Overall accuracy of the local-storm environment (i.e., within a 600-km wide, WRF-centered box) and large-scale environment (i.e., entire model domain) was evaluated using the dry energy norm (Rabier et al. 1996):

$$\langle \mathbf{X}, \mathbf{Y} \rangle = -\frac{1}{2} \int_{p_{sfc}}^{p_{top}} \iint_A \left(\Delta u^2 + \Delta v^2 + R_d T \Delta \ln(p_{sfc})^2 + \frac{c_p}{T_v} \Delta T^2 \right) dx dy dp \quad (1)$$

In (1), \mathbf{X} is the WRF model state vector, \mathbf{Y} is the GMA state vector, u is the zonal wind (m s^{-1}), v is the meridional wind (m s^{-1}), R_d is the dry air gas constant ($287 \text{ J kg}^{-1} \text{ K}^{-1}$), T_r is the mean surface temperature (K), P_{sfc} is the surface pressure (Pa), c_p is heat capacity at constant pressure ($1004 \text{ J kg}^{-1} \text{ K}^{-1}$), and T is air temperature (K). A 600-km wide box was used to evaluate the local-storm environment because it captured the storm, yet minimized background environment contamination. Point-to-point root mean square error (RMSE) calculations were performed for sea-level pressure (SLP), 850-hPa temperature, 500-hPa geopotential height, 300-hPa winds, and 2 potential vorticity unit (2 PVU; $1 \text{ PVU} = 10^{-6} \text{ K m}^2 \text{ kg}^{-1} \text{ s}^{-1}$) potential temperature. The first four variables were frequently referenced in Kocin and Uccellini (2004) for nor'easter analysis, and 2-PVU potential temperature was used to investigate changes to the dynamic tropopause. For both the energy norm and RMSE, smaller values denote less error.

The energy norm and RMSE metrics served complimentary purposes. First note that the energy norm is a volume integration, whereas RMSE is a layer integration. Thus the former better represents the entire model simulation and was less sensitive to large errors in single layers. Second, the energy norm involves four variables (surface pressure, temperature, zonal wind, and meridional wind) and not just one, making its results more robust. Buizza et al. (2005) provided a compelling argument for the use of the energy norm as the primary validation metric and its current usage at ECMWF for model

validation. Given all the above, the energy norm was the primary vehicle to evaluate model version simulation quality, and RMSE helped to identify error sources at levels throughout the troposphere.

2.3. Results

2.3.1. Comparison to Cavallo et al. (2010) local potential temperature tendencies

Due to the RRTM emphasis in this paper, differences in composite six-hour, local potential temperature tendencies between GFS and WRF model forecasts were calculated for comparison to Cavallo et al. (2010). Figures 1c and 1d show this quantity for both Cavallo et al. (2010) and for the present work, respectively. Cavallo et al. (2010) based their composite upon six-hour model predictions from WRF 3.1.1 model forecasts initialized every six hours during the time period starting 00 UTC 10 August 2009 and ending 06 UTC 30 August 2009 using a single 36 km resolution model domain (Fig. 1a). Comparatively, the Fig. 1d composite was derived from six-hour model forecasts from domain 1 of each case (Fig. 1b). In addition to the local potential temperature tendency (light grey, dashed line) Fig. 1c also depicts the longwave radiative potential temperature heating rate (black, solid line) and its one standard deviation limit (shading) as well as the net longwave potential temperature heating rates for their entire model domain. It is from these heating rates that the -15K day^{-1} cold bias claimed by Cavallo et al. (2010) was based. For reference, standard longwave heating rates are shown for the mid-latitude summer (MLS; medium grey solid line) and tropics (TROP; medium grey, dashed line).

Comparing Figs. 1c and 1d revealed W311 local potential temperature tendencies composites exhibited similar characteristics, but they were not completely identical.

Similar to Cavallo et al. (2010) both composites demonstrate a weak cold bias (at least 0.25K) below 900hPa, and a slightly warming bias ($\sim 0.5 \text{ K day}^{-1}$) both between 500-850 hPa and around the tropopause at 200 hPa. Between 50-100 hPa, both composites exhibit either a neutral or in Cavallo et al. (2010) a slight cooling bias ($\sim 0.25 \text{ K day}^{-1}$). One missing element from Fig. 1d was a significant cooling bias (up to 10 K day^{-1}) above 50 hPa because the model top in the present study was exactly 50hPa. Other smaller discrepancies are likely related to the smaller number of six hour forecasts (8 vs 121 samples), the different months and years, and the smaller spatial coverage in this study. Thus the shown heating rates in Fig. 1d would be particularly sensitive to large-scale weather patterns on the eight days sampled.

All WRF model run composites shown in Fig. 1d exhibit characteristics (sign and levels) similar to that seen in W311 with the exception of W32. For the W32 profile, its shape is similar to W311, except it exhibited a cold bias relative to the GFS six-hour forecast. Analysis of 700-hPa potential temperatures (not shown) revealed this cold bias did not greatly alter simulated dynamical fields as compared to other WRF model runs. Despite the W32 cold bias, WRF model simulations with the newer RRTM parameterization did demonstrate a reduced cold bias above 100hPa consistent with Cavallo et al. (2010).

2.3.2 WRF-version analysis

Figure 2 depicts SLP-based storm tracks for all WRF simulations. In Fig. 2 track error relative to GMA ranged from relatively minor (cases 3 and 5) to extreme (cases 2 and 4). To quantify WRF track variability and ascertain potential biases, Fig. 3 displays GMA-relative track errors every six hours (small symbols) and their mean (larger

symbols) from each WRF run. These errors are not latitude/longitude-based, but instead relative to the GMA cyclone propagation direction every six hours which defines the positive y-direction. Using this framework, simulation track bias ranged from 50 km (case 5) to over 2,300 km (case 4), with a typical value of 150-300 km. Overall, 20 out of 32 (62.5%) WRF simulations exhibited a leftward track bias averaging 75 km and 20 out of 32 (62.5%) simulations lagged GMA averaging 75 km. To further quantify these biases Table 2 indicates the total number of six-hour periods where each individual model simulation exhibited a particular track bias relative to GMA. As seen in Table 2, at least 58% of six hour periods exhibited a leftward bias or lagged GMA. Inter-WRF comparisons revealed W311 had an equal probability of either a leftward or rightward track bias and WRF33 had at least 88 km additional track error on average versus other WRF simulations.

To investigate larger simulated storm track errors, as well as inter-WRF track error variability, a more in-depth analysis of cases 2 and 4 was completed. Figure 4 displays various dynamical fields for GMA, W311 and W33 for case 2 at 18UTC on 16 March 2007. This time was selected because each WRF simulation had produced a cyclone and the differences relative to GMA were apparent. For case 2, all model simulations generated the initial surface low too far north by over 400 km (Fig. 2) resulting in an overall leftward model track bias (Fig. 3). As seen in Figs. 4a-c, the WRF simulated 300-hPa jet centered over Ontario has a similar magnitude, alignment, and position to GMA. Because actual cyclogenesis occurred over Alabama and Georgia, well-removed from the jet, it was eliminated as a possible candidate to explain the observed track error in case 2. Instead 500-hPa vorticity advection was partly to blame

(Figs. 4d-f). In GMA, a positive vorticity maximum ($\sim 12 \times 10^5 \text{ s}^{-1}$) is located near Charleston, South Carolina where the surface cyclone was located. Given the southwestward thermal wind at 500-hPa (not shown), upward vertical motion can be inferred across eastern North Carolina via the Trenberth approximation (Trenberth 1978). In comparison, both W311 (Fig. 4e) and W33 (Fig. 4f) failed to generate the surface cyclone because both lacked any significant positive vorticity advection in this same region. Instead the combination of upward vertical motion attributed to the 300-hPa jet and strong mid-tropospheric latent heat release (not shown) may have helped lead WRF astray. Additionally, boundary conditions in the study were derived from GFS model forecasts which lead to the possibility of GFS model errors propagating into WRF simulations and decreasing their accuracy.

Figure 5 displays various dynamical fields at two key times from case 4. These two times were selected to emphasize storm initialization and its later maturation. As seen in Figs. 5a-f, fields of 500-hPa geopotential height closely match GMA on 28 February but 48 hours later they varied radically from GMA. Eventually, accumulated model error resulted in a 72-hour time lag with an over 1800 km (Figs. 2 and 3) track error between WRF and GMA. The source of this gargantuan track error was a persistent 500-hPa geopotential height cut-off trough situated over the southeastern US (Figs. 5d-f). While both GMA and WRF initially produced this cut-off, GMA maintained it for roughly one day, whereas WRF maintained it for over four days. Due to this cut-off, nor'easter propagation speed fell sharply to under 5 m s^{-1} in each WRF run, at the same time GMA propagated it rapidly northeastward, which led to the significant track error in case 4 (Fig. 3).

In addition to large track error relative to GMA, case 4 WRF model simulations also demonstrated larger variability than in other cases. Inter-WRF track differences in case 4 can be attributed to positional and intensity variations of simulated mid-tropospheric latent heat release. As seen in Figs. 5 g-l, on 28 February simulated SLP and latent heat release had only subtle differences between WRF runs (e.g., latent heat release off New Jersey). As mentioned above, all WRF simulations two days later mistakenly maintained a 500-hPa cut-off height minimum (Figs. 5d-f), but despite their similarities, the SLP fields and location of latent heat release varied strikingly between W33 and the other WRF runs. For SLP this difference was most apparent for the 1008 hPa contour on 2 March (Figs. 5 j-l). Both W311 and W32 extended this contour from North Carolina to east of Jacksonville, FL, but W33 extended it well into the Bahamas. The larger extent of the 1008 hPa contour in W33 is likely linked to the more southern location of higher latent heat release as compared to W311 and W32.

To assess simulated cyclone intensity accuracy, minimum SLP was compared between each WRF run to GMA. Simulated SLP was deemed “accurate” if it was within 5 hPa of GMA. Overall, WRF model simulations over-intensified a nor’easter upwards of 16.9 hPa (case 1, W33) and under-intensified by as much as 20.5 hPa (case 6; W311-MRad), which equates to 2.20 and -1.77 times the standard deviations of 9.34 hPa and 9.56 hPa for W33 and W311-MRad, respectively. Individually, no WRF version simulated cyclone intensity accurately in a majority of the eight cases. Specifically, such simulations were accurate in 1, 1, 3 and 2 of 8 cases for W311, W311-MRad, W32, and W33, respectively. With the exception of W33, no WRF version exhibited a distinct over-intensification or under-intensification bias. For W33, it over-intensified in 4 out of 8

cases, averaging 9.5 hPa, under-intensified in 2 out of 8 cases, averaging 8.1 hPa, and was accurate in 2 out of 8 cases. When over-intensification occurred in other WRF simulations, it was often of smaller magnitude (average: 6.19 hPa) than W33. However, when under-intensification occurred in non-W33 simulations, the bias was larger than in W33 (average: 9.46 hPa versus 8.0 hPa).

Plots of key meteorological fields (e.g., temperature, relative vorticity, latent heating, upper-level jet divergence, etc.) were generated to investigate why simulated storm intensity varied. Analysis of these data found SLP to be most strongly associated with variations in mid-tropospheric latent heat release. To illustrate, Fig. 6 displays 500-hPa latent heating (Fig. 6a-d) and 300-hPa winds (Fig. 6e-h) with SLP overlaid at both 12 and 18 UTC on 15 October 2009. These were selected because simulated cyclones were within 50 km relative to each other, cyclone intensity at 12UTC was within 1 hPa, and synoptic-scale fields were similar. Given all the above, SLP variations were more easily attributed to particular fields. At 12 UTC both W32 and W33 have strong latent heat release ($>20 \text{ K day}^{-1}$) northeast of the cyclone center which itself was in the right-exit region of the 300-hPa jet. Six hours later, minimum SLP from W32 fell 3.6 hPa and W33 fell 2.81 hPa. At 18UTC latent heating remained strong along the cyclone track and both simulated cyclones remained in the right entrance region of the 300-hPa jet. Local changes in geopotential height, 850-hPa temperatures, and 500-hPa relative vorticity (not shown) were comparatively small. Further observation of Figs. 6a-d revealed latent heating was stronger ($\sim 5 \text{ K day}^{-1}$) near the cyclone center in W32 as compared to W33, resulting in enhanced upward vertical motion and larger pressure falls in W32. For all other cases, latent heat release intensity was linked to SLP falls, but unlike in Fig. 6 it

acted in concert with other favorable factors (e.g., positive vorticity advection) to produce the observed inter-WRF SLP variability.

Local-storm environment simulation accuracy was evaluated via energy norm differences (W311 - model) for both GMA- and WRF-centered frameworks (Fig. 7). The energy norm in each framework was evaluated within a 600-km-wide box which was centered on GMA and each WRF cyclone, respectively. In Fig. 7, thin lines represent GMA-centered and thick lines represent WRF-centered energy norm differences where positive values denote improvement versus W311. Table 3 indicates the number of six-hour periods where W311-MRad, W32, and W33 runs had a lower energy norm than WRF311 in each framework. For comparison, the number of six hour periods where W32 had a lower energy norm than W33 is also included. Overall, using newer versions of WRF improved up to 29 out of 63 (46.0%) WRF-centered periods, and 30 out of 63 (47.6%) GMA-centered periods. For individual models, WRF-centered energy norm differences were positive in more than 50% of the periods in 5, 3, and 2 of 8 cases for W311-MRad, W32, and W33, respectively. This latter result is potentially misleading due to the high case-to-case period variability which ranged from 5 to 11 total time periods per case and because in 9 out of 24 comparisons W311 did better than other WRF versions by only one additional period. Energy norms from W33 were lower than W32 in 28 out of 63 (44.4%) WRF-centered periods and 4 out of 8 cases overall. These results suggest that the newer RRTM parameterization and W32 updates did not significantly alter simulation outcome given that energy norm differences for W311-MRad and W32 are near zero. On the other hand, W33 saw much higher variability (4x) than either W311-MRad or W32. Specifically, on a few occasions (case 1, 00 UTC 23 Nov.; case 4,

18 UTC 2 March) W33 had a notably lower energy norm than W311, but a majority of periods clearly show increased energy norms. Potential explanations for the W33 results include possible updated parameterization conflicts, or unforeseen conflicts related to the WRF dynamical core update, or changes to the default WRF parameterization options or compiler flags. To investigate these possibilities however would require running W33 with many additional parameterization configurations.

To identify potential energy norm sources, WRF- and GMA-centered energy norms were time-averaged and horizontally-integrated at each vertical level (Fig. 8). In Fig. 8, the following four energy norm spikes were common to all cases: ~850 hPa, ~600 hPa, ~300 hPa, and ~180 hPa. Further analysis (not shown) attributed these energy norm spikes to errors in the planetary boundary layer height and the low-level jet, mid-tropospheric jet intensity and positioning, upper-level jet stream intensity and positioning, and tropopause height, respectively. Despite changes associated with the RRTM scheme, W311 energy norms were not always higher above 100 hPa than in other WRF simulations as might be expected from Cavallo et al. (2010). The lack of increased above 100 hPa is possibly explained due to both the energy norm in Fig. 8 being time-averaged (not just at six hours) and because it measures the error between WRF and GMA and not between WRF and GFS forecasts. Figure 8 also reveals the largest energy norm contributor in both the GMA- and WRF-relative environments was at 650 hPa and is driven by errors in simulated winds (not shown). Analysis of mid-tropospheric winds revealed a jet that was in close proximity to each nor'easter in both WRF forecasts and GMA. This jet in WRF simulations was often misaligned relative to GMA, resulting in higher model error. To illustrate this misalignment, Fig. 9 displays 500-hPa winds from

both W32 and W33 (grey lines) overlaid on GMA (black lines) during cases 3 and 7. Both WRF runs simulate the 500-hPa jet, but it is displaced by more than 100 km relative to GMA, and misshapen.

Energy norm differences integrated over the entirety of domains 1, 2 and 3 were used to assess the broader impacts from each WRF version. To illustrate how energy norm differences varied with model domain, Fig. 10 depicts these differences for cases 5 (Figs. 10 a,c,e) and 8 (Figs. 10 b,d,e). As seen in Fig. 10, the energy norms from W311-MRad and W32 are nearly identical to W311 given their small associated energy norm differences. Comparatively, W33 energy norm differences were up to an order of magnitude greater than other model runs (Fig. 10a) and largely detrimental to nor'easter simulation accuracy. Further analysis of energy norm components revealed wind-related errors to be the largest energy norm contributor. As an example, in Fig. 9c-d depict a strong 50 m s^{-1} (100 kt) jet at 500 hPa in GMA (black lines). Both models (grey lines) displace this jet too far westward, but the position, extent, and magnitude of the 500-hPa jet streak in W32 is noticeably closer to GMA than W33. Principal energy norm sources were commensurate with the local-storm environment analysis, except domain 1 energy norms around at 600 hPa and 300 hPa were sharply decreased and increased, respectively. Increased energy norm at 300 hPa was attributed to the large spatial coverage (approximately 1,000-3,000 km long and 500-1000 km wide) and intensity (up to 75 m s^{-1}) of the jet (not shown). For domains 2 and 3, the 300-hPa jet was not present during all time periods and it covered a lower percentage of grid points as compared to domain 1. In contrast the mid-tropospheric jet is of lower magnitude ($<40 \text{ m s}^{-1}$) and smaller spatial extent (500-1000 km long and 300-500 km wide) than the 300-hPa jet, but

its intensity appears linked to the nor'easter. Because of these attributes and its association with the nor'easter, the mid-tropospheric jet had a modest error contribution on domain 1, yet larger contributions for domains 2 and 3.

As compared to W311, nor'easter simulations exhibited lower energy norms in newer model versions for at most 184 (74.2%), 162 (65.3%), and 143 (57.7%) out of 248 periods for domains 1, 2 and 3, respectively (Table 3). Unlike the local-storm analysis, the results in Table 3 reflect each case and time period having equal weight so that the results are more definitive. Across all eight cases, W311-MRad had a lower energy norm than W311 on average in 6, 6, and 5 cases for domains 1, 2, and 3, respectively. Because W311-MRad improved only up to 142 out of 248 (57.3%) time periods and energy norm differences were near zero, these increases in simulation accuracy relative to W311 were modest. For W32, energy norms on average were lower than W311 in 8, 8, and 6 cases for domains 1, 2, and 3, respectively. The superior simulation performance of W32 relative to W311 led up to 184 out of 248 (74.2%) time periods being improved. Similar to W311 however, energy norm differences, although larger, were still close to zero suggesting only modest simulation accuracy gains compared to W311. Finally, W33 on average had lower energy norm than W311 in 2, 2, and 3 cases for domains 1, 2, and 3, respectively. The worse accuracy of W33 is supported by Table 3 which shows the number of time periods improved by W33 over W311 was nearly 100 fewer (domain 1) than W32. As previously illustrated by Figs. 6 and 9, increased error in W33 is at least partially attributed to errors in simulated winds, mid-tropospheric latent heat release and other dynamical fields.

For comparison to the energy norm results and to determine which meteorological fields were most sensitive to WRF model version, RMSE calculations were completed and summarized in Table 3. The five RMSE error variables shown in Table 3 were selected either due to their strong emphasis in Kocin and Uccellini (2004) or as an indicator of the dynamic tropopause accuracy. Similar to the energy norm, values shown in Table 3 reflect the number of six-hour periods where a particular model has a lower RMSE value than W311. The last column shows this same data, except that it compares W32 to W33. Results from the five RMSE calculations were generally consistent with the energy norm results and indicated W32 produced the best simulations overall. To illustrate this consistency, Fig. 11 displays energy norm (Fig. 11a) and five RMSEs (Figs. 11 b-e) from case 4. Of the eight cases, case four was selected because it best exemplified the trends seen in other cases. As shown in Fig. 11, RMSE differences generally mirrored the energy norm with the exception of 2 PVU potential temperature. This exception occurred in all WRF model runs. Possible explanations include errors originating with wind and temperature fields in the upper troposphere which in-turn alter potential vorticity values, gravity wave damping near the model top, or the high degree of fine-scale variability common to 2 PVU potential temperature fields. Root mean square error results from W311-MRad indicate the inclusion of RRTM 3.2 led to a 50% or more time period improvement in 10 out of 15 (66.7%) RMSE variable evaluations. Figure 11 however indicated that simulation improvement relative to W311 was modest. For W33, only 6 out 15 (40%) RMSE variables evaluations had 50% or more of all time periods improved over W311. As shown in Fig. 11, RMSE differences between W33 and W311

were up to an order of magnitude larger than other WRF versions and typically indicated increased error with W33.

2.4. Conclusions

Nor'easter simulation accuracy from three WRF model releases (3.1.1, 3.2, and 3.3) and one modified WRF 3.1.1 simulation using RRTM 3.2 were evaluated. All model simulations were configured identically, run for 180 model hours, initialized 72 hours prior to first precipitation impacts in the highly populated Mid-Atlantic US, and evaluated using GMA. Simulated cyclone track accuracy varied from case to case with track errors ranging from 50 (case 5) to 2,300 km (case 4), with 200-300 km more typical. A majority of WRF simulations (20 out of 32) favored a leftward bias and lagged GMA, but track errors associated with W33 tended to be larger (~80 km). These biases were attributed to errors from simulated fields of relative vorticity, geopotential height, and mid-tropospheric latent heat release and potentially GFS-forecast-based boundary conditions. Accurate model-simulated nor'easter intensity (± 5 hPa) was not achieved in a majority (best: W32 3 out of 8 cases) of the cases for any WRF version. Only W33 exhibited a definitive over-intensification bias which averaged approximately 9.5 hPa. All other model versions were equally likely to either over-intensify or under-intensify a cyclone. However, when under-intensification did occur in non-W33 runs, the cyclone was weakened on average an additional 1.4 hPa as compared to W33. Further analysis revealed these cyclone intensity differences were attributed to variations in simulated mid-tropospheric latent heat release.

Both WRF- and GMA-centered energy norm differences proved an invaluable tool to evaluate nor'easter simulation accuracy. Non-W311 simulations improved up to 29 out of 63 (46.0%) WRF-centered periods, and 30 out of 63 (47.6%) GMA-centered periods versus W311. Simulations from W311-MRad and W32 did not demonstrate a tangible improvement in simulation accuracy as evidenced by their near zero energy norm differences. Energy norms from W33 had four times the variability other WRF runs, and were typically higher versus W311. Time-averaged energy norm profiles revealed four distinct energy norm spikes associated with the planetary boundary layer height and the low-level jet, mid-tropospheric jet intensity and positioning, upper-level jet stream intensity and positioning, and tropopause height. Notably, energy norm contributions from winds were often the greatest, especially at 650 and 300 hPa, which were associated with errors in jet streak position and intensity.

Energy norm and RMSE calculations across the entirety of domains 1, 2, and 3 provided a broader picture as to how WRF version impacted the entire simulation. Energy norm results for W311-MRad and W32 showed improvement in more than 50% of all time steps for 142 out of 248 and 184 out of 248 periods, respectively. Despite such numbers, energy norm differences were relatively close the zero which suggests only modest gains in simulation accuracy versus W311. With the exception of 2 PVU potential temperature, RMSE results were commensurate with the energy norm findings. The inconsistency of 2 PVU potential temperature RMSE with other results was noted to be likely associated with gravity wave damping related errors and the high degree of fine-scale variability commonly associated with this and other meteorological parameters (Marzban and Sandgathe 2006; Elbert 2008). Energy norm and RMSE differences from

W33 were an order of magnitude higher than other WRF runs, but these differences were only positive for as few as 86 out of 248 time periods for the energy norm, denoting a general degradation of simulation accuracy. The poor showing of W33 was shown to be at least partly associated to errors in simulated winds, mid-tropospheric latent heat release and other dynamical fields and possibly differences in the default WRF configuration and compilation options.

Overall, this study has shown that newer model releases are not always more accurate than their predecessors. While the inclusion of more complex parameterizations and model physics are necessary as NWPM simulations reach finer resolutions and continue to evolve, the modeling community must not lose sight of the ultimate objective of numerical weather prediction: Accurate weather simulations. Also it is important to recall that each component and program that comprised a complex NWPM such as WRF is in actuality only piece within a larger complex puzzle. This last point is best illustrated by the RRTM update. In Cavallo et al. (2010) the longwave radiative heating rate cold bias was reduced from over 15 K day^{-1} to nearly zero, an impressive result. Despite such improvement, however this update failed to produce any notable changes to simulation accuracy as compared to W311.

2.5. Future work

This work broadly addressed how WRF model version impacted nor'easter simulations. Results raised from this work leave much to be explored in future follow-up studies. In the present work, W33 produced less accurate simulations than both W311 and W32. This work attributed the W33 performance to potential complications related to updated WRF parameterizations, default configuration options, or the WRF dynamical

core update. Future studies could run additional simulations with different parameterizations to ascertain whether the results of this study have any parameterization dependence. Another potential research route would be to include the recently released WRF model version 3.4. Running WRF 3.4 simulations would allow investigation of whether the worse performance of W33 was strictly limited to that release. Next, a nor'easter is only one of a multitude of extreme weather events. Future work could potentially investigate how different WRF model versions fared during other winter weather events (e.g., polar lows), tropical cyclones, or perhaps simulations of convection along the inter-tropical convergence zone. Finally, for this study GMA served as ground truth for model validation, future studies could compare WRF model performance at various weather stations in the US or compare variables such as cloud cover extent and snow cover to satellite data from sources such as the Moderate Resolution Imaging Spectroradiometer. Additionally, given the challenges in verifying high-resolution forecasts against lower resolution ground truth data, additional verification methods such as object-oriented (Marzban and Sandgathe 2006) or fuzzy verification (Elbert 2008) could be utilized.

Table 1: Nor'easter case list. Column 3 shows the dates which each event impacted the northeastern US, while the last two columns denote the first and last times for each model run.

Case Number	NESIS	Event Dates	Model Run Start Date	Model Run End Date
1	N/A	22-24 Nov. 2006	11/19 12UTC	11/27 00UTC
2	2.54	15-17 Mar. 2007	3/12 18UTC	3/20 06UTC
3	N/A	15-17 Apr. 2007	4/12 06UTC	4/19 18UTC
4	1.65	1-2 Mar. 2009	2/26 12UTC	3/6 00UTC
5	N/A	15-16 Oct 2009	10/12 12UTC	10/20 00UTC
6	4.03	19-20 Dec. 2009	12/16 06UTC	12/23 18UTC
7	4.38	4-7 Feb. 2010	2/2 18UTC	2/10 06UTC
8	N/A	12-14 Mar. 2010	3/9 06UTC	3/16 18UTC

Table 2: GMA-relative storm track bias analysis. Values indicate the total the number six-hour time periods from all cases where each model exhibited the shown track bias relative to GMA at each time period.

	Left		Right		Key	
Lead	17	20	14	17		
	16	21	17	14	W311	W311 MRad
Lag	29	32	19	13	W32	W33
	33	35	17	19		

Table 3: Performance analysis relative to W311. Values denote the number and percentage of 6-hour periods where the energy norm or RMSE from the indicated simulations was lower than W32. The last section indicates the number of periods where COSMIC-based energy norms were lower than W32 within a 600 km box centered on GMA and each simulated storm, respectively. Bolded values indicate the best WRF simulation (as compared to W311) for each analysis component.

Domain 1 Performance Analysis vs W311 (248 Total Times)	W311-MRad	W32	W33	W32 < W33
Energy Norm	131 (52.8%)	184 (74.2%)	89 (35.9%)	171 (69.0%)
RMSE 2 PVU Pot T	161 (64.9%)	174 (70.2%)	112 (45.2%)	165 (66.5%)
RMSE 300-hPa Winds	115 (46.4%)	164 (66.1%)	86 (34.7%)	169 (68.1%)
RMSE 500-hPa Geo Hght	129 (52.0%)	146 (58.9%)	139 (56.0%)	121 (48.8%)
RMSE 850-hPa Temps	155 (62.5%)	158 (63.7%)	134 (54.0%)	124 (50.0%)
RMSE SLP	150 (60.5%)	109 (44.4%)	109 (44.0%)	137 (55.2%)
Domain 2 Performance Analysis vs W311 (248 Total Times)	W311-MRad	W32	W33	W32 < W33
Energy Norm	142 (57.3%)	162 (65.3%)	86 (34.7%)	170 (68.5%)
RMSE 2 PVU Pot T	127 (51.2%)	156 (62.9%)	123 (49.6%)	140 (56.5%)
RMSE 300-hPa Winds	118 (47.6%)	139 (56.0%)	108 (43.5%)	137 (55.2%)
RMSE 500-hPa Geo Hght	121 (48.8%)	142 (57.3%)	129 (52.0%)	127 (51.2%)
RMSE 850-hPa Temps	154 (61.1%)	165 (66.5%)	118 (47.6%)	135 (54.4%)
RMSE SLP	120 (48.4%)	128 (51.6%)	102 (41.1%)	144 (58.1%)
Domain 3 Performance Analysis vs W311 (248 Total Times)	W311-MRad	W32	W33	W32 < W33
Energy Norm	132 (53.2%)	143 (57.7%)	121 (48.8%)	131 (52.8%)
RMSE 2 PVU Pot T	128 (51.6%)	149 (60.1%)	135 (54.4%)	123 (49.6%)
RMSE 300-hPa Winds	113 (45.6%)	111 (44.8%)	125 (50.4%)	121 (48.8%)
RMSE 500-hPa Geo Hght	110 (44.4%)	150 (60.5%)	136 (54.8%)	115 (46.4%)
RMSE 850-hPa Temps	136 (54.8%)	158 (63.7%)	122 (49.2%)	135 (54.4%)
RMSE SLP	133 (53.6%)	134 (54.0%)	112 (45.2%)	146 (58.9%)
Domain 2 Storm Performance Analysis vs W311 (63 Total Times)	W311-MRad	W32	W33	W32 < W33
Energy Norm GFS Rel	30 (47.6%)	30 (47.6%)	25 (39.7%)	35 (55.6%)
Energy Norm WRF Rel	29 (46.0%)	28 (44.4%)	26 (41.3%)	36 (57.1%)

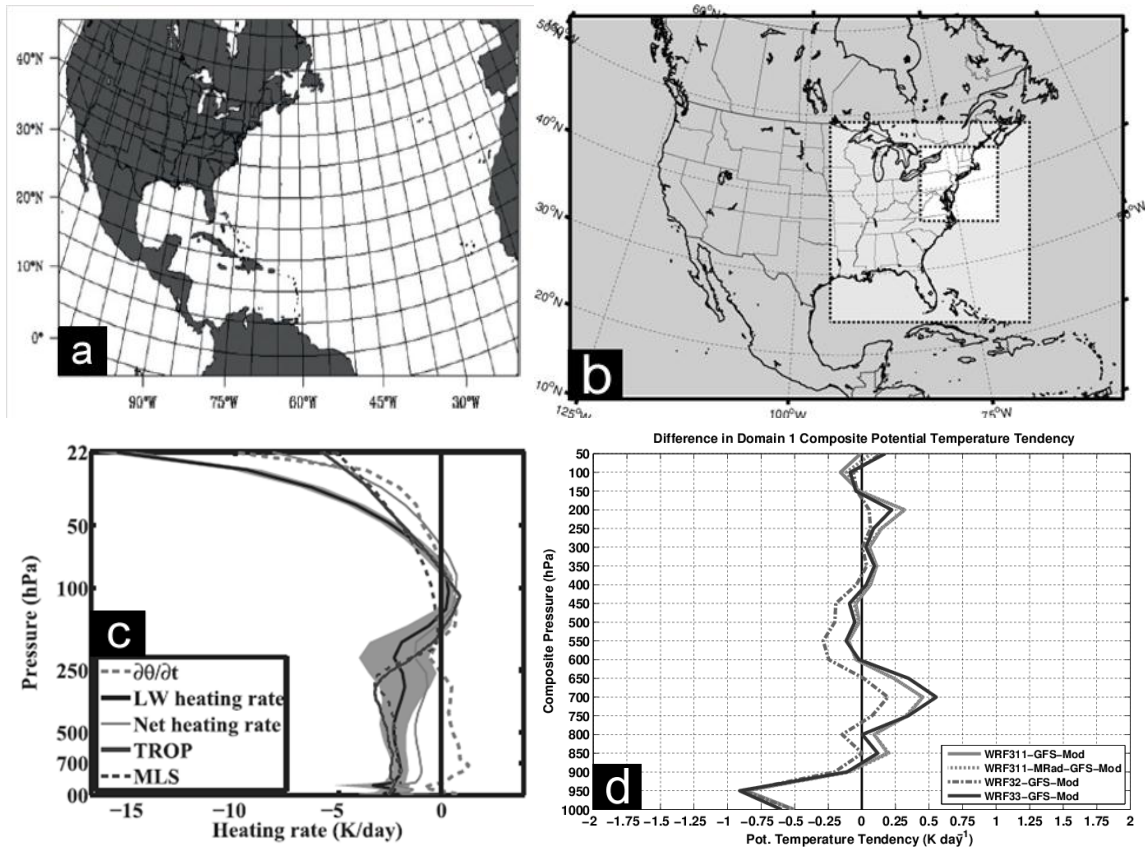


Fig. 1: Comparison of model domain and difference in composited, six-hour potential temperature tendency forecasts between WRF and the GFS. Panels a and b depict the model grid configuration from Cavallo et al. (2010; their Fig 1a) and the present study, respectively. Panel c shows various heating rates from Cavallo et al. (2010; their Fig 1b); potential temperature tendency is shown as the right-most, dotted line. Panel d shows the eight-case composite of potential temperature tendency from each WRF model version for domain 1.

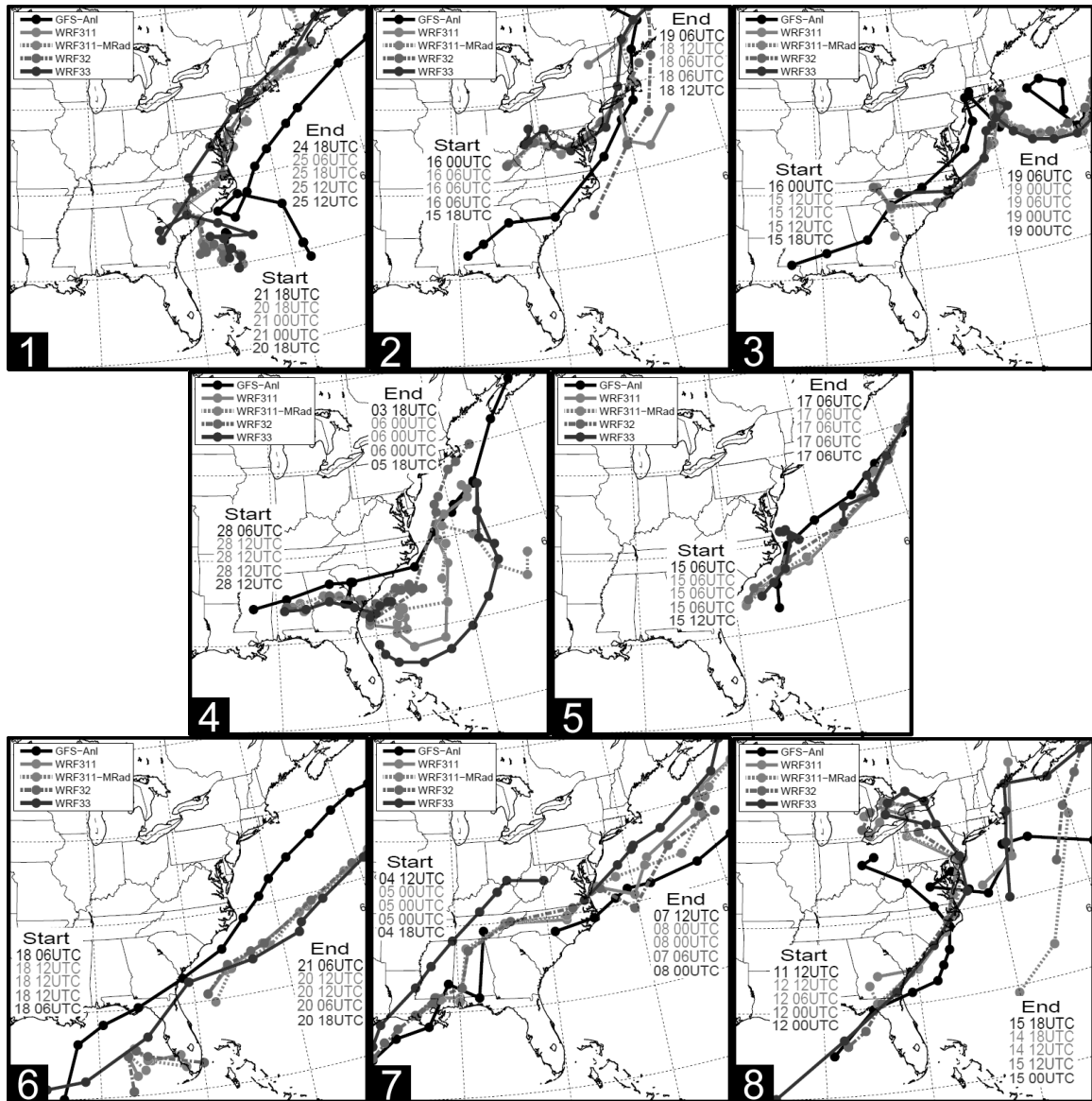


Fig. 2: Storm tracks from GMA and each WRF version run. Line legend is shown on the upper-left of each plot. Shown symbols indicate simulated storm position every six hours. White numbers indicate case number.

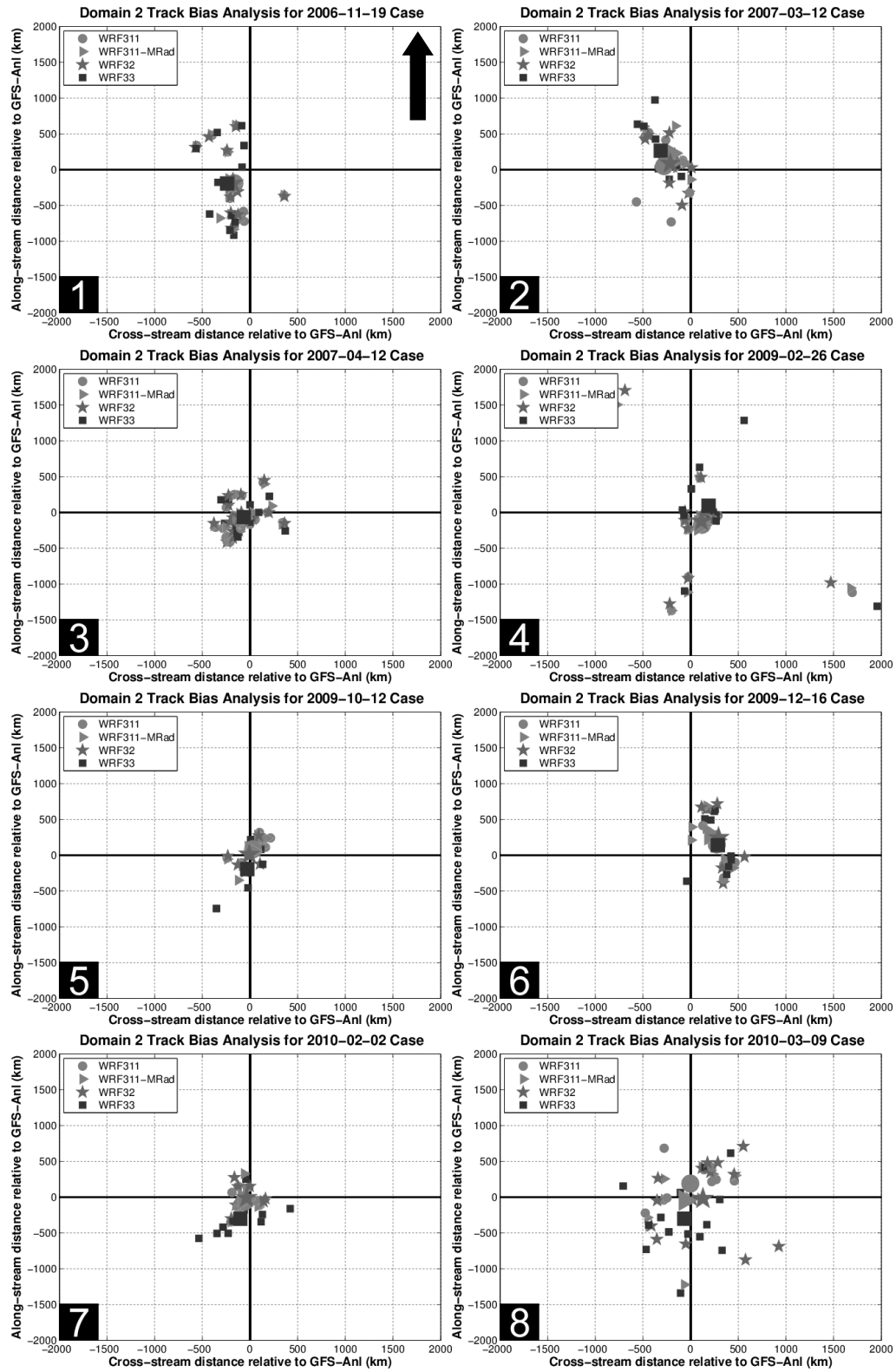


Fig. 3: WRF forecasted storm position bias as compared to GMA for all eight cases. The black arrow in panel 1 indicates the GMA storm motion direction for all panels. Shown symbols represent WRF position bias every six hours (smaller symbols) and their mean (large symbols).

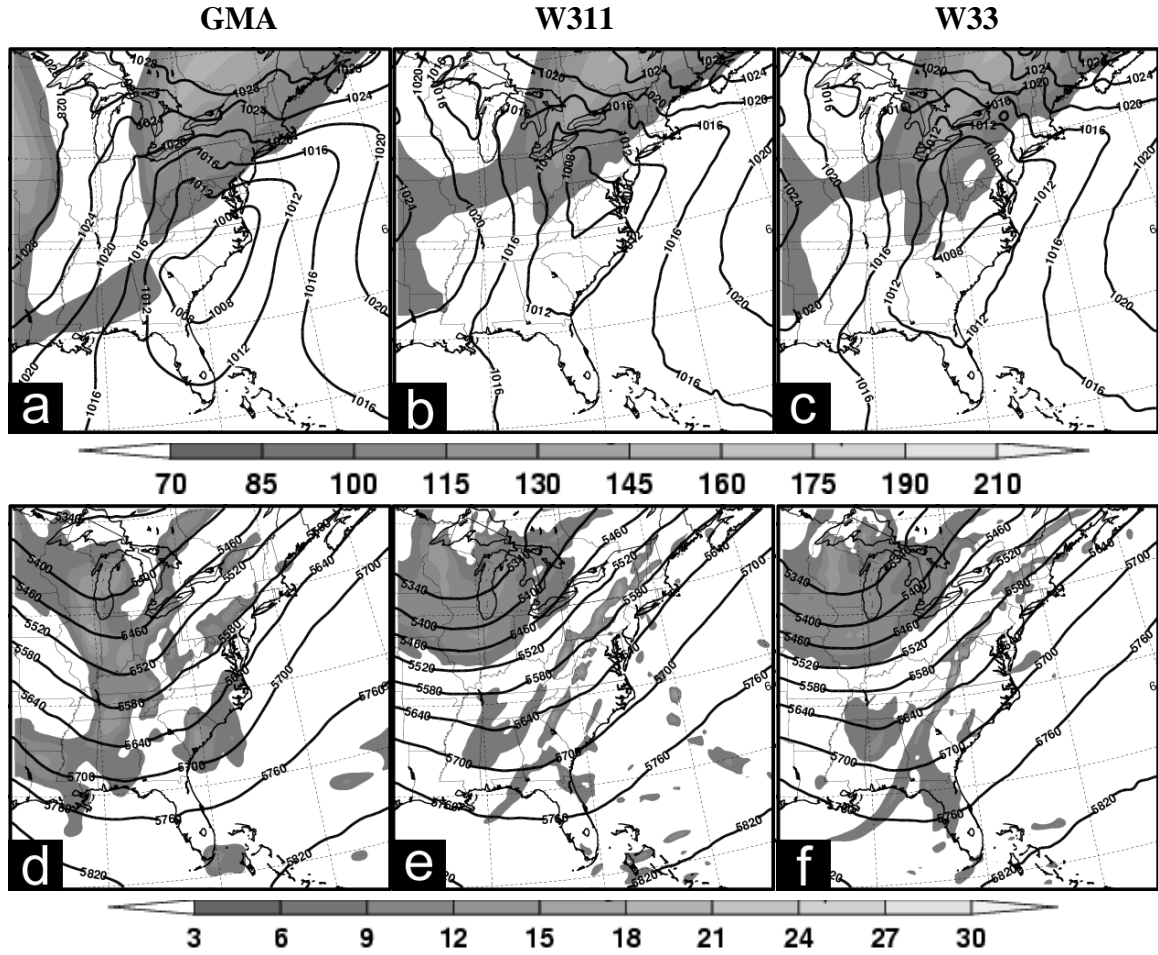


Fig. 4: Various plots from (a,d) GMA, (b,e) W311, and (c,f) W33 (c, f). (a-c) 300-hPa wind speed (fills, knots) and SLP (contours, hPa) and (d-f) 500-hPa positive relative vorticity (fills, 10^{-5} s^{-1}) and 500-hPa geopotential height (contours, m) on 18 UTC 16 March 2007.

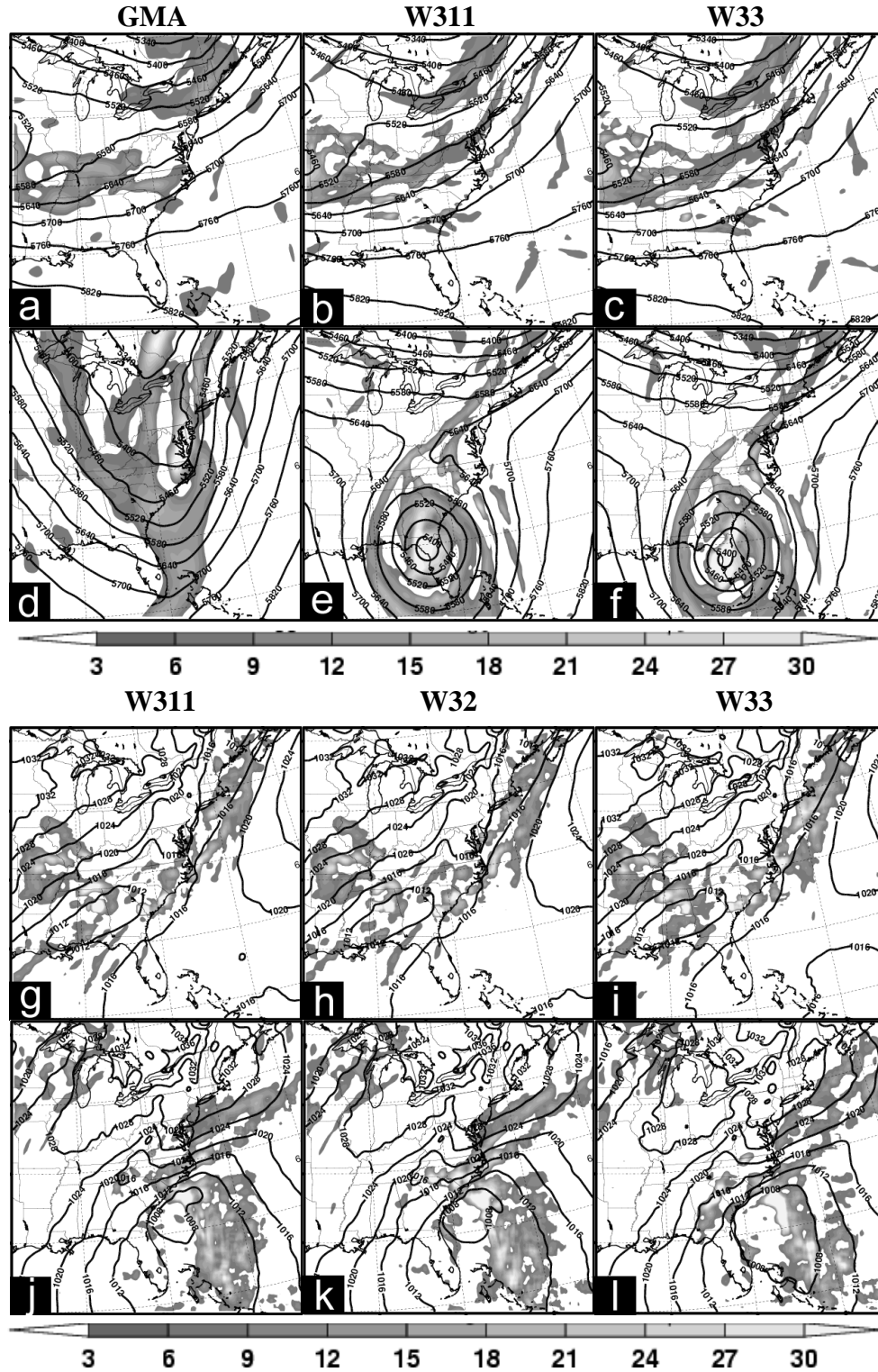


Fig. 5: Various plots from (a,d) GMA, (b,e,g,i) W311, (h,k) W32 (h, k) and (c,f,i,l) W33. (a-f) 500-hPa positive relative vorticity (fills, 10^{-5} s^{-1}) and 500-hPa geopotential height (fills, m) on 28 February 2009 12UTC (a-c) and 12 UTC 2 March 2009 (d-f). (g-l) 500-hPa latent heating (fills, K day^{-1}) and sea-level pressure (contours, hPa) on 28 February 2009 12UTC (g-i) and 12 UTC 2 March 2009 (j-l).

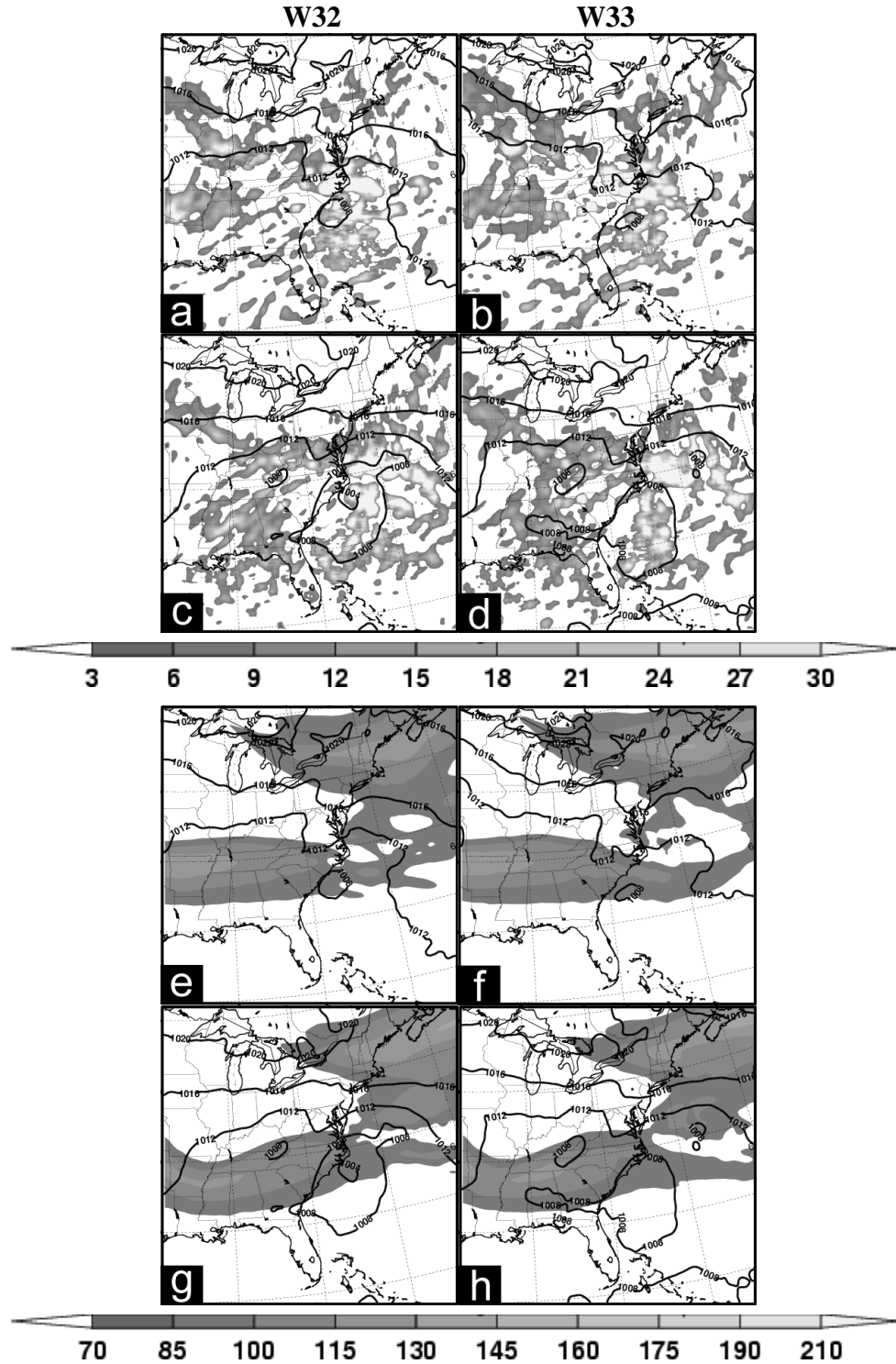


Fig. 6: (a-d) 500-hPa latent heating (fills, K day⁻¹) and sea-level pressure (contours, hPa) and (e-h) 300-hPa winds (fills, knots) and sea-level pressure (contours, hPa). Upper panels (a, b, e, f) are from 15 October 2009 at 12UTC and the lower panels (c, d, g, h) are from 18UTC the same day.

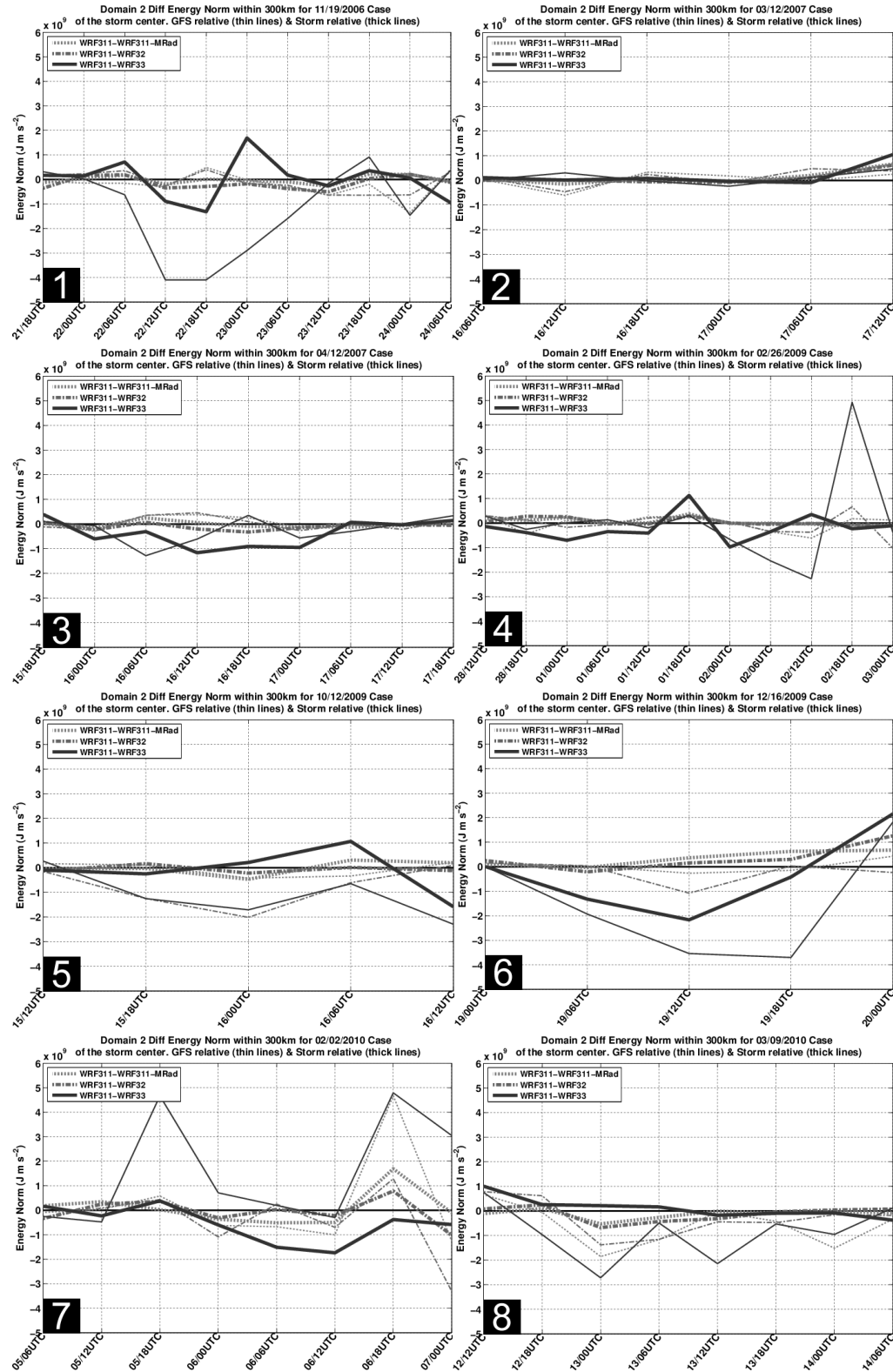


Fig. 7: Domain 2 differences in energy norms for all cases within 300 km of the GMA storm center (thin lines) and each model simulated center (thick lines). Shown differences are relative to WRF 3.1.1. Positive values denote improvement over WRF 3.1.1.

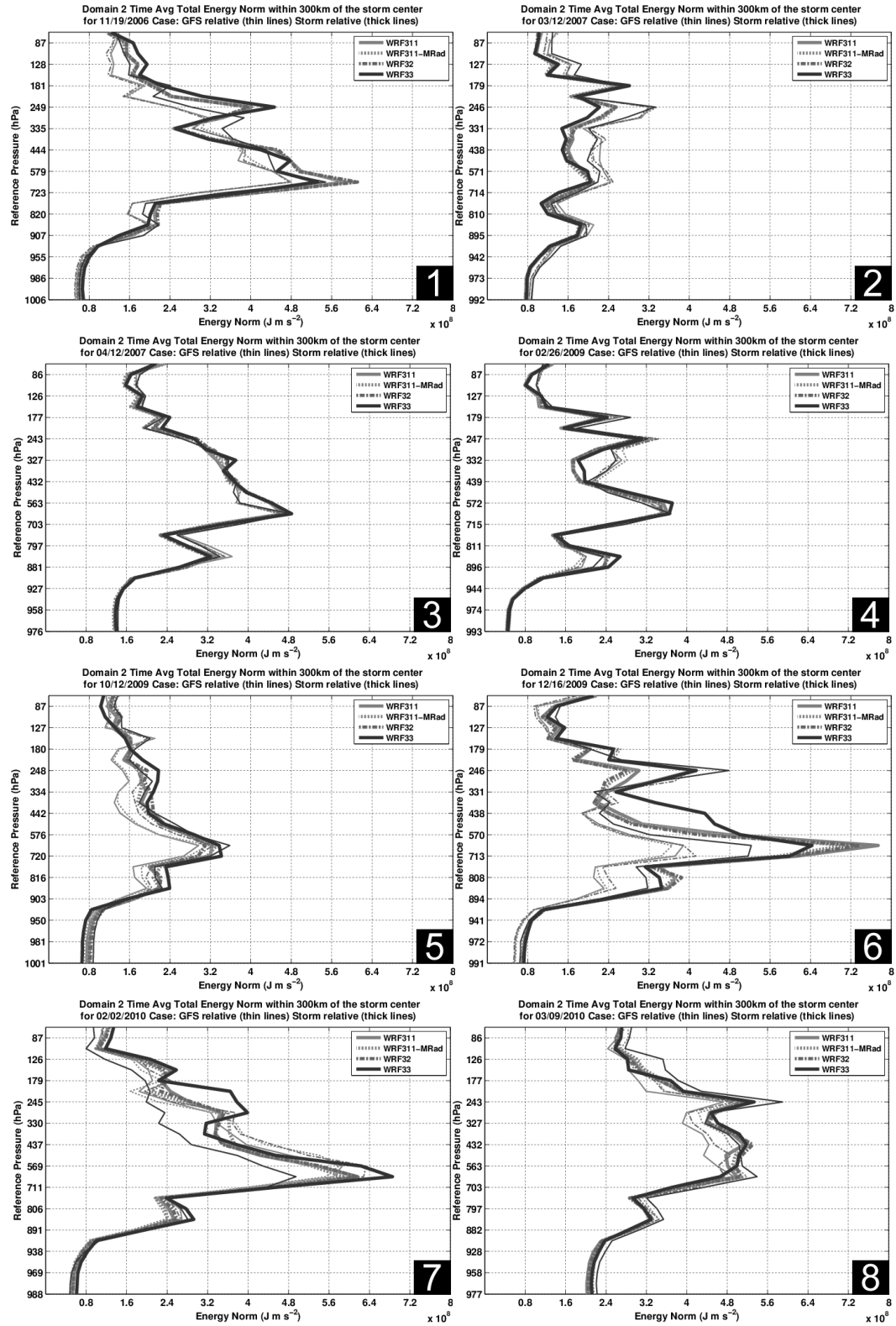


Fig. 8: Domain 2, time-averaged energy norm for all cases within 300 km of the GFS analysis storm center (thin lines) and each model simulated center (thick lines).

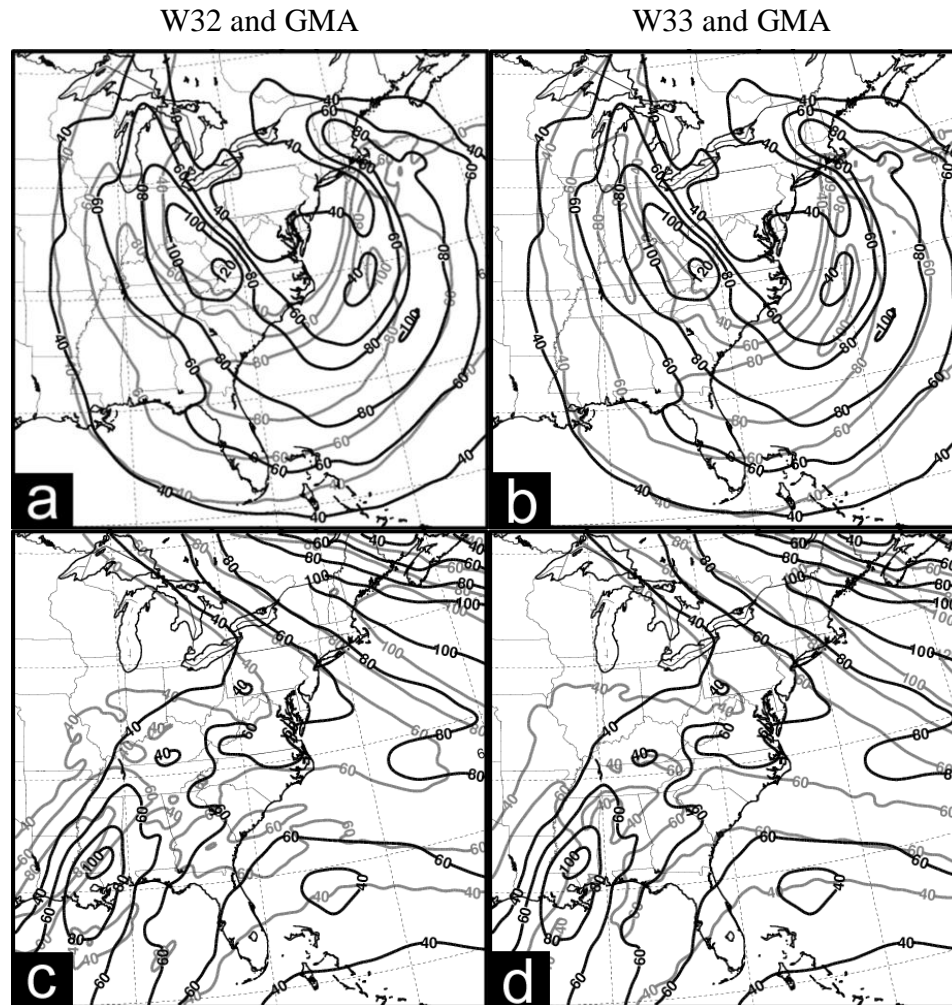


Fig. 9: 500-hPa wind speed plotted for (a, c) W311 (grey) and GMA (black) and (b, d) W32 (grey) and GMA (black). (a, b) are on 12 UTC 16 April 2007 and (c,d) are for 12 UTC 5 February 2010.

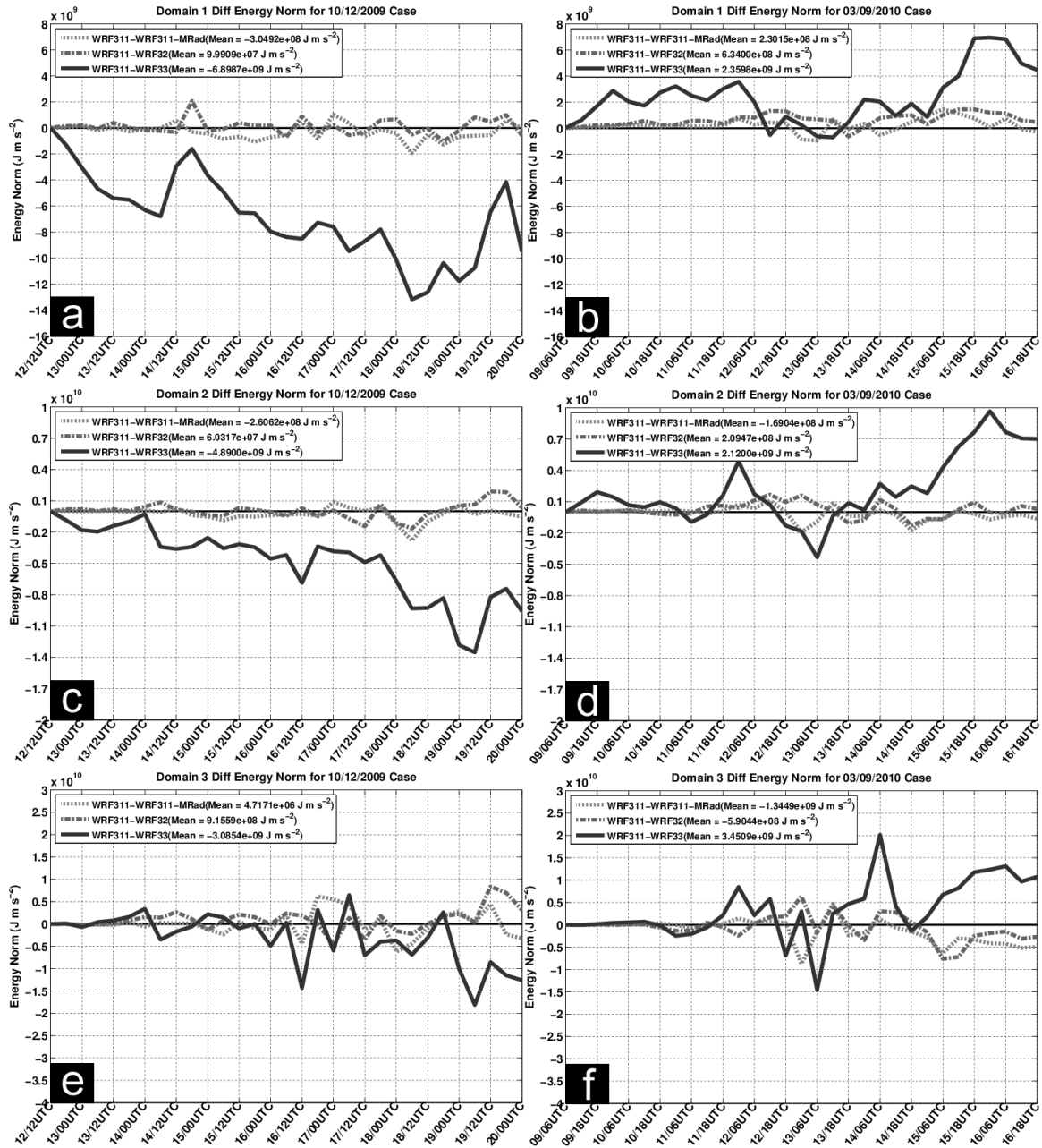


Fig. 10: Energy norm differences within the entirety of domains 1, 2, and 3 for (a, c, e) case 5 and (b, d, e) case 8. Shown differences are relative to WRF 3.1.1. Positive values denote improvement over WRF 3.1.1.

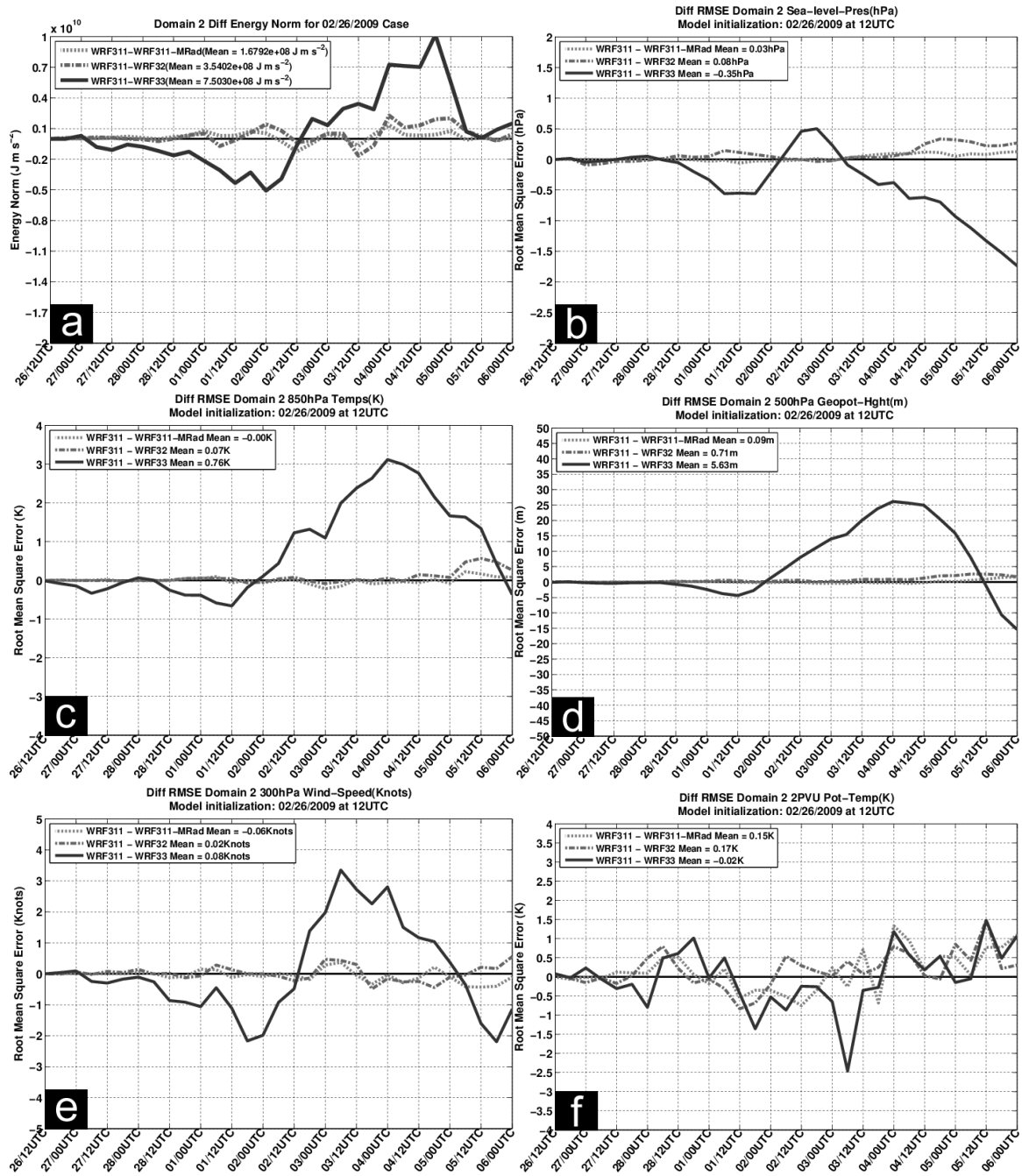


Fig. 11: Difference in domain 2 (a) energy norm, (b) sea-level pressure, (c) 850-hPa temperature, (d) 500-geopotential height, (e) 300-hPa winds, (f) 2-PVU potential temperature for case 4. Shown differences are relative to WRF 3.1.1. Positive values denote improvement over WRF 3.1.1.

3. Influence of microphysics schemes upon numerical simulations of nor'easters

3.1 Chapter abstract

The impact of three, five-class and four, six-class bulk microphysics schemes (BMPS), created between 1983-2011, on Weather Research and Forecasting Model simulations of eight nor'easter simulations was assessed. Model simulations were conducted for 180 hours, starting roughly 72 hours prior to the first precipitation impacts in the highly populated Mid-Atlantic US and associated cyclogenesis. Simulation accuracy was assessed (except hydrometeor mixing ratios) by comparing each to Global Forecasting System model analysis.

Overall, BMPS WRF simulations exhibited notable storm track errors (up to 2,200 km) and intensity differences (up to 15.6 hPa) versus model analysis. Between model runs however, no BMPS run demonstrated consistent improvement in either storm track or intensity relative to other schemes. Due to their common programming heritage, simulated liquid hydrometeors (cloud ice and rain) varied negligibly between BMPSs. On the other hand, frozen hydrometeor species (cloud ice, graupel, and snow) varied considerably between BMPS due to their various assumptions concerning hydrometeor concentrations and terminal velocities. Simulations of the local storm environment were most accurate for the most recent five- and six-class BMPS. On the other hand, simulations of over the entirety of four nested domains revealed the newest BMPS scheme only performed on-par with the oldest BMPS within the inner most model domains.

3.2. Methods

3.2.1. Study design

We utilized the W33 which solves fully-compressible, non-hydrostatic, Eulerian equations in terrain-following coordinates (Skamarock et al. 2008; Wang et al. 2012). To simulate a multi-scale phenomenon such as a nor'easter, model grids of sufficient size and resolution were required. The three-domain grid (Fig. 11) had two-way interaction, 45, 15, 5, and 1.667 km grid spacing, respectively, 61 vertical levels, and a 50-hPa (~20 km) model top. The spatial extent of the outer 3 domains afforded detection of both synoptic (e.g., upper-level jet, longwave troughs) and mesoscale (e.g., shortwave troughs, mid-level latent heat release) phenomena, whereas the inner-most domain was of sufficiently high resolution to illustrate differences between BMPs. There were 61 vertical levels for consistency with past and present WRF model studies at NASA-Goddard Space Flight Center (i.e., Shi et al. 2010; Tao et al. 2011). Boundary conditions were derived from Global Forecasting System (GFS) model forecasts ($1^\circ \times 1^\circ$ resolution).

Model simulations were conducted for 180 hours, starting roughly 72 hours prior to the first precipitation impacts in the highly populated Mid-Atlantic US and associated cyclogenesis. This time frame focused attention exclusively on cyclone initiation and its later impact in this region. A 72-hour lead time allowed simulations to spin-up, establish baroclinicity between the cooler eastern United States and warmer Gulf Stream, and simulate latent heating along the expansive (>1000 km) northern edge of the Gulf Stream. All of the above are vital for accurate nor'easter simulations (Kuo et al. 1991; Mote et al.

1997; Yao et al. 2008). Precipitation data from the New Jersey Weather and Climate Network (Robinson 2005) served as a proxy for establishing when each nor'easter first impacted the Mid-Atlantic US. Using these data, model initialization was set as 72 hours prior to the first nor'easter-related 0.5 mm (~ 0.02 inch) precipitation reading. A New Jersey-centric approach was chosen because of its high population density ($461.6/\text{km}^2$), significant contribution (\$473 billion) to the US gross domestic product, and its relatively central location in the region of interest (United States Census Bureau 2012).

The selection of WRF model parameterizations conforms to past and on-going WRF studies at Goddard Space Flight Center (i.e., Shi et al. 2010; Tao et al. 2011).

- Longwave radiation: New Goddard Scheme (Chou and Suarez 2001)
- Shortwave radiation: 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: Grell-Devenyi ensemble scheme (Grell and Devenyi 2002)

The BMPSs used in this study and their corresponding citations are summarized in Table 4. For this work, we used three, five-class (WRF Single Moment, five-class scheme [WSM5] and two versions of the Goddard Cumulus Ensemble [GCE]-2ice) and four, six-class (Lin, two versions of the GCE-3ice-graupel and WSM6) BMPSs. Unlike previous studies, we used both the newest, unreleased version of the GCE (Lang et al. 2011) and the most recent WRF release (Lang et al. 2007). To differentiate the 2007 and 2011 GCE BMPS releases, hereafter they will be referred to as the Old Goddard

(hereafter OGod) and New Goddard (hereafter NGod) schemes, respectfully. In all, 7 total simulations were completed for each nor'easter case.

This study focused on eight nor'easter cases (Table 5) selected based upon their timing, intensity, and track. All cases occurred after an April 2006 satellite launch to correlate with a parallel nor'easter study (Nicholls and Decker 2012) and during the “nor'easter season” (October to April) as defined by Jacobs et al. (2005). To make this study more generalized, we did not select only the most severe nor'easters, but instead aimed for a relatively small but diverse sample. Case severity was measured using the Northeast Snowfall Impact Scale (NESIS; Kocin and Uccellini 2004). Under NESIS, nor'easters were classified on a scale of 1 (notable) to 5 (extreme) based upon the population impacted, area affected, and event severity. Finally, all cases were required to have precipitated in New Jersey as mentioned above.

3.2.2. Verification and analysis techniques

Validation data was derived from GFS model analysis (GMA). This data source was chosen because all model domains included data sparse regions where in-situ observations were frequently unavailable, and GMA was easily interpolated to the WRF model grid. Furthermore, the lateral boundary conditions were GFS-based.

Model run analysis was comprised of several parts. Storm tracks were determined using local-minima in SLP via an objective, self-developed, WRF-track algorithm similar to that used at the Climate Prediction Center (Serreze 1995; Serreze et al. 1997). To elucidate storm track differences and potential biases each simulation was compared to GMA every six hours. Then, the hydrometeor mixing ratios produced from each BMPS

run were inter-compared. Similar to WP10, the hydrometeor analysis was qualitative, because GMA only contains water vapor mixing ratio estimates. Overall accuracy of the local-storm environment (i.e., within a 600-km wide, WRF-centered box) and large-scale environment (i.e., entire model domain) was evaluated using the dry energy norm (Rabier et al. 1996):

$$\langle \mathbf{X}, \mathbf{Y} \rangle = -\frac{1}{2} \int_{P_{sfc}}^{P_{top}} \iint_A \left(\Delta u^2 + \Delta v^2 + R_d T (\Delta \ln(p_{sfc}))^2 + \frac{c_p}{T_r} \Delta T^2 \right) dx dy dp \quad (1)$$

In (1), \mathbf{X} is the WRF model state vector, \mathbf{Y} is the GMA state vector, u is the zonal wind (m s^{-1}), v is the meridional wind (m s^{-1}), R_d is the dry air gas constant ($287 \text{ J kg}^{-1} \text{ K}^{-1}$), T_r is the mean surface temperature (K), P_{sfc} is the surface pressure (Pa), c_p is heat capacity at constant pressure ($1004 \text{ J kg}^{-1} \text{ K}^{-1}$), and T is air temperature (K). A 600-km wide box was used to evaluate the local-storm environment because it captured the storm, yet minimized background environment contamination. Point-to-point root mean square error (RMSE) calculations were performed for sea-level pressure (SLP), 850-hPa temperature, 500-hPa geopotential height, 300-hPa winds, and 2 potential vorticity unit (2 PVU; $1 \text{ PVU} = 10^{-6} \text{ K m}^2 \text{ kg}^{-1} \text{ s}^{-1}$) potential temperature. The first four variables were frequently referenced in Kocin and Uccellini (2004) for nor'easter analysis, and 2-PVU potential temperature was used to investigate changes to the dynamic tropopause. For both the energy norm and RMSE, smaller values denote less error.

The energy norm and RMSE metrics served complimentary purposes. First note that the energy norm is a volume integration, whereas RMSE is a layer integration. Thus the former better represents the entire model simulation and was less sensitive to large errors in single layers. Second, the energy norm involves four variables (surface pressure, temperature, zonal wind, and meridional wind) and not just one, making its results more

robust. Buizza et al. (2005) provided a compelling argument for the use of the energy norm as the primary validation metric and its current usage at ECMWF for model validation. Given all the above, the energy norm was the primary vehicle to evaluate model version simulation quality, and RMSE helped to identify error sources at levels throughout the troposphere.

Model results will be segregated into two parts: Five-class and six-class microphysics schemes. These two BMPSS types are separated to both prevent cluttering figures and tables with too much information, but more importantly all six-class BMPSS include graupel, whereas the five-class BMPSS did not. Thus schemes using the latter would be at an inherent disadvantage because graupel exists in the real atmosphere. Despite this disadvantage, general comparisons between the five and six-class BMPSS will be made to evaluate the inclusion of graupel may have impacted simulations.

3.3. Results

3.3.1. Five-class microphysical schemes

Figure 13 depicts SLP-based storm tracks for all WRF simulations and GMA. In Fig. 13 track error relative to GMA ranged from relatively minor (cases 3 and 5) to extreme (cases 2 and 4). To quantify WRF track variability and ascertain potential biases, Fig. 14 displays GMA-relative track errors every six hours (small symbols) and their mean (larger symbols) from each WRF run. These errors are not latitude/longitude-based, but instead relative to the GMA cyclone propagation direction every six hours which defines the positive y-direction. Using this framework, simulation track bias ranged from 50 km (case 5) to over 2,300 km (case 4), and typically between 240-250 km. The lack of

significant inter-WRF track variability in Fig. 13 and the small difference (10 km) in average track error are both consistent with WP10 and Tao et al. (2011). Overall, 17 out of 24 (70.8%) WRF simulations exhibited a leftward track bias averaging 81 km and 14 out of 24 (58.3%) simulations lagged GMA averaging 66 km. To further quantify these biases Table 6 indicates the total number of six-hour periods where each individual model simulation exhibited a particular track bias relative to GMA. As seen in Table 6, at least 53.3% (WSM5) and as many as 61.1% (OGoD) of six hour periods exhibited a leftward bias or lagged GMA. These biases suggest that each WRF simulations may generate each nor'easter either too late or propagate it too slowly.

To investigate larger simulated storm track errors, as well as inter-WRF track error variability, a more in-depth analysis of cases 2 and 4 was completed. Figures 15a-c display 500-hPa relative vorticity and geopotential height for GMA, WSM5 and NGoD for case 2 on 18UTC 16 March 2007. As seen in Fig. 13, all model simulations initiated surface cyclogenesis 400-800 km too far northward as compared to GMA. As illustrated by Fig. 15, this error is largely explained by the lack of strong positive vorticity maximum ($>21 \times 10^{-5} \text{ s}^{-1}$) over southern Alabama where the GMA surface cyclone generated. Each WRF simulation instead generated the surface cyclone near a wide region of 500-hPa latent heat release near the Great Lakes and close to the right-entrance region of the 300-hPa jet (not shown). Further WRF simulation errors may have arisen from the GFS-based boundary condition errors.

Figures 15e-f display 500-hPa relative vorticity and geopotential height for GMA, WSM5 and NGoD for case 4 on 06UTC 2 March 2009. As seen in these figures, the two WRF model simulations simulated a cut-off 500-hPa geopotential height minimum over

southeast Georgia. In contrast, GMA shows this same cut-off height minimum starting to merge into poleward longwave trough at a location roughly 600 km further northward than in any WRF model run. All WRF simulations slowed the cyclone to a lethargic (5 m s^{-1}), whereas in GMA the 500-hPa cutoff rapidly propagated northeastward and fully merged with the poleward shortwave height trough. Accumulated model errors in WRF eventually produced a 72-hour time lag with an over 1800 km (Figs. 13 and 14) track error between WRF and GMA. Further analysis demonstrated that all WRF simulations lacked sufficient positive vorticity advection (Figs. 15e-f) as compared to GMA (Fig. 15d), which eventually caused the WRF simulated cyclones to become vertically stacked.

Next, to assess simulated cyclone intensity accuracy, minimum SLP was compared between each WRF run and GMA. Simulated SLP was deemed “accurate” if it was within 5 hPa of GMA. Overall, model simulations over-intensified a nor’easter upwards of 15.6 hPa (case 1; WSM5) and under-intensified by as much as 9.5 hPa (case 3; WSM5), which equates to 1.26 and -1.57 times the standard deviations of 7.41 hPa. All WRF simulations accurately predicted SLP in cases 6 and 7, over-intensified case 1, and under-intensified case 3. For the other four cases, cyclone intensity was accurate in 2 out of 3 WRF simulations for each case. For these four cases, all three simulations for each case only showed minor differences in meteorological fields (i.e., SLP, 500-hPa latent heating, 500-hPa geopotential heights) and no BMPS was consistently worse than the other two competing schemes. The only minor variations between the simulations was not unexpected given that all three five-class BMPS schemes are all derived from the Lin BMPS scheme.

We next focused our attention on how four hydrometeorological species (cloud water, cloud ice, rain, and snow) varied amongst the five-class BMPSs using column-integrated mixing ratios (Fig. 16) and composited, time-averaged profiles (Fig. 17). For consistency with WP10 and Tao et al. (2011), we only analyzed domain 4 due to its 1.667-km resolution. In Figure 16, column-integrated mixing ratios are shown on 06 UTC 6 Feb 2010. Although other times were comparable, this particular time was selected because all WRF simulations exhibited small variation in storm track (<50 km) and had each microphysical species present in domain 4. Similar to WP10, both column-integrated and composite cloud water and rain mixing ratios profiles vary modestly. Our composite profiles (Fig. 17) show nearly identical maximum saturation heights, and cloud water and rain mixing ratio concentrations that are not distinctly different. As explained in WP10, this similarity is likely because both the Goddard schemes and WSM5 BMPS codes are derived from Lin.

Despite their common heritage in Lin, both GCE and WSM5 address frozen hydrometeors (snow and cloud ice) differently. For snow mixing ratios, all BMPS schemes simulate a similar coverage area, but WSM5 has noticeably smaller values consistent with Tao et al. (2011). Smaller WSM5 snow mixing ratios can be attributed to the different snow hydrometeor size distribution intercept parameters as well as to the interactions between snow and other hydrometeors via auto conversion and accretion. We found the OGod scheme produced up to twice as much snow as NGod which corresponds well to Lang et al. (2011). In Lang et al. (2011), increased snow in OGod is caused by its high snow growth bias from other ice microphysical processes (especially riming) and its higher snow distribution intercept parameter. The composite snow mixing ratio profiles

produced similar results. In Figure 17b, the pressure level of maximum snow saturation does not vary notably between the three BMPSs, but the maximum concentrations are smallest in WSM5 and largest in OGod for the reasons mentioned earlier.

In comparison to snow mixing ratios, there was significant variability among five-class BMPS in column-integrated cloud ice (Fig. 16). Figure 16 shows WSM5 generated appreciably more cloud ice than either Goddard BMPS. These high cloud ice concentrations shown Fig. 16 are associated with notably larger cloud ice concentrations below 500 hPa in WSM5 (Fig. 17c). The root cause of this excess cloud glaciation in WSM5 was an unintentional consequence of the overhauled Lin BMPS ice microphysics utilized in WSM5 (WP10). This modification successfully addressed a known excess cloud ice generation bias at cold temperatures in the Lin BMPS, but sharply increased cloud ice generation at temperatures between 0°C and -20°C (WP10). Despite this known issue with WSM5, at mid and upper levels OGod and WSM5 are nearly indistinguishable but different from the NGod scheme. The NGod has less cloud ice overall, and its maximum value is reached at a much lower pressure which Lang et al. (2011) attributed to its temperature dependent cloud-ice size distribution functions. Cloud-ice size distribution functions in the other BMPS are fixed.

Local-storm environment simulation accuracy was evaluated via energy norm differences (WSM5 - model) for both GMA- and WRF-centered frameworks (Fig. 18). The energy norm in each framework was evaluated within a 600-km-wide box which was centered on GMA and each WRF cyclone, respectively. In Fig. 18, thin lines represent GMA-centered and thick lines represent WRF-centered energy norm differences where positive values denote improvement versus WSM5. The bottom of Table 7 displays the

number of six-hour periods where the OGod and NGod runs had a lower energy norm than WSM5 in each framework. In comparison to WSM5, the OGod and NGod runs had less simulation error in 26 (44.1%) and 34 (57.6%) out of 59 WRF-centered periods, respectively. At GMA-centered locations these two BMPS had smaller error than WSM5 in 20 (33.9%) and 42 (71.2%) out of 59 periods. For individual models, WRF-centered energy norm differences were positive in more than 50% of the periods in 4 and 5 of 8 cases for OGod, and NGod, respectively. Comparing the two Goddard BMPSs revealed that the NGod scheme frequently had less error (41 out of 59 time periods) in either framework. These results suggest NGod produced the most accurate simulations of the local-storm environment. Given the common Lin heritage between all three, five-class BMPSs suggests a possible importance of a temperature-dependent snow intercept which was unique to the NGod scheme. This result however may merit some caution given that the small energy norm differences in Fig. 18 and that some cases only had five time periods compared.

To identify potential energy norm sources, WRF- and GMA-centered energy norms were time-averaged and horizontally-integrated at each vertical level (Fig. 19). In Fig. 19, three energy norm spikes were common to all cases: ~900 hPa, ~750 hPa, and ~200 hPa. Further analysis (not shown) attributed these energy norm spikes to errors in the planetary boundary layer height and the low-level jet, mid-tropospheric jet intensity and positioning, and tropopause height, respectively. Figure 19 also reveals the largest energy norm contributor in both the GMA- and WRF-relative environments was at 650 hPa and is driven by errors in simulated winds (not shown). Analysis of mid-tropospheric winds revealed a jet that was in close proximity to each nor'easter in both WRF forecasts

and GMA. This jet in WRF simulations was often misaligned relative to GMA, resulting in higher model error.

Energy norm differences integrated over the entirety of domains 1, 2, 3 and 4 were used to demonstrate how each of the BMPSs impacted overall simulation accuracy. To illustrate how this quantity varied Fig. 20 depicts these differences for cases 3 (Figs. 20a, 20c, 20e, 20g) and 8 (Figs. 20b, 20d, 20f, 20h). As seen in Fig. 20, the NGod scheme is both fairly consistent (same patterns between domains) and fairs well against both WSM5 and OGod. Because the OGod is the direct predecessor to the NGod scheme the appearance of its energy norm differences is fairly similar to the NGod scheme. Unlike the NGod scheme however, OGod simulations were more typically on-par or slightly worse than WSM5, but more importantly it has more error on coarser domains. The latter result suggests a potential grid spacing dependence of the OGod scheme with respect to simulations accuracy. Time-averaged and horizontally-integrated energy norms over the entire domain (not shown) were generally consistent with the local-storm analysis, except in addition to model error also increased around 300 hPa in association with the upper tropospheric jet stream. Between domains the contribution to the energy norm from errors with the 300-hPa jet decreased with model domain due to its decreasing influence within each smaller domain. On the other hand, the impact of the 600-hPa jet increased with model domain due to its association with each nor'easter case and its direct impact within each model domain.

As compared to WSM5, OGod and NGod simulations had smaller energy norms in at most 127 (51.2%; NGod), 137 (55.2%; NGod), 135 (54.4%; NGod), and 122 (49.2%; NGod) out of 248 six-hour periods for domains 1, 2, 3 and 4, respectively (Table

7). For domains 1, 2, 3, and 4, OGod had smaller model error than WSM5 in 1, 0, 1, and 4 out of 8 cases, respectively and in as many as 118 (47.6%) out of 248 periods. Simulations from NGod exhibited a lower energy norm than WSM5 in 4, 4, 5, and 3 out of 8 cases, respectively and in as many as 137 (55.2%) out of 248 periods. Direct comparison between OGod and NGod revealed the latter had a lower energy norm in all except domain 4 in up 208 (83.8%; domain 2) of 248 periods and for 7, 8, and 5 out of 4 cases, respectively. Therefore, the energy norm results suggest NGod to be slightly more accurate than either WSM5 or the OGod. Given that energy norm differences, with only a few exceptions (domain 3, case 8), were often 2 orders of magnitude less than the energy norm itself suggest only small changes to overall simulation accuracy.

For comparison to the energy norm results and to determine which meteorological fields were most sensitive to WRF model version, RMSE calculations were completed and summarized in Table 7. The five RMSE error variables shown in Table 7 were selected either due to their strong emphasis in Kocin and Uccellini (2004) or as an indicator of the dynamic tropopause accuracy. Similar to the energy norm, values shown in Table 7 denote the number of periods where each BMPS had a lower RMSE than WSM5. As seen in Table 7, the RMSE results were not always consistent with those from the energy norm. Specifically, for all four model domains, RMSE results from 300-hPa winds and 500-hPa geopotential height contradicted the energy norm results. To illustrate this observation graphically, Fig. 21 displays the energy norm (Fig. 21a) and five RMSEs (Figs. 21 b-e) from case 6 from domain 2. This particular case was selected because it is representative of trends seen in the other seven cases. As shown in Fig. 21, the contradiction between the energy norm and the two RMSEs is profound with NGod

exhibiting more error relative to WSM5 than the OGod scheme in nearly every time step. A potential explanation for the contradictory results at 500 and 300 hPa are the temperature dependent snow intercept and cloud ice size distribution. As seen in Fig. 17 the these temperature dependent additions in the NGod scheme raised the level of maximum cloud ice (from 500-hPa to 280 hPa) and decreased the maximum snow mixing ratio as compared to OGod at around 500 hPa. Changes to these concentration may impact latent heat release in the mid and upper troposphere and thus impact both winds and geopotential heights. Levels above 280 hPa and below 500-hPa show little variation. As suggested in Fig. 21, the size of the RMSE differences relative to the meteorological parameters upon which they are based can be quite small (> 2 orders of magnitude). Although should be noted that domain 4 300-hPa wind RMSEs were sometimes within 1 order of magnitude of the average wind speed.

To summarize the five-class scheme results using the most comprehensive metric, the energy norm, the NGod scheme marginally outperforms both WSM5 (51.2% or more improved periods) and OGod (60.1% or more improved period) for domains 1-3. For domain 4, WSM5 slightly outperformed (50.8% of periods improved) both Goddard schemes. Second, combining all 24 tests (energy norm and RMSE from all domains), OGod exhibited the lowest model error in 10 out of 24 tests, and both WSM5 and NGod were lowest in 7 out of 24 tests. Finally, with respect to the local storm environment, the energy norm results indicated NGod better simulated the nor'easter in 57.6% and 69.8% of all periods than the WSM5 and OGod runs, respectively

3.3.2. Six-class microphysical schemes

Figure 22 depicts SLP-based storm tracks for all six-class BMPS simulations and GMA. In Fig. 22 track error relative to GMA ranged from relatively minor (cases 3 and 5) to extreme (cases 2 and 4). To quantify WRF track variability and ascertain potential biases, Fig. 23 shows GMA-relative track errors from the six-class BMPS runs ranged from 25 km (case 5) to over 2,200 km (case 4), and typically between 255-305 km. Similar to WP10 and the five-class BMPSs the average inter-WRF track variability was a modest 50 km. In comparison to five-class BMPSs, both Goddard schemes tracked nearly identical, however WSM6 storm tracks did vary from WSM5 (most notably cases 2 and 8). Overall, 24 out of 32 (75.0%) six-class simulations exhibited a leftward track bias averaging 113 km and 22 out of 32 (68.7%) simulations lagged GMA averaging 81 km. To further quantify these biases, Table 8 shows the number of six-hour period where each simulation type exhibited a particular track bias relative to GMA for all cases. With the exception of WSM6, all six-class simulations demonstrated a leftward track bias for 63.3% (NGod) to 66.7% (OGod) of the time, yet all simulations tended to lag GMA between 53.8% (Lin) to 62.2% (NGod) of the time. In comparison to the five-class schemes, WSM6, OGod, and NGod was favored either a leftward bias or lagged GMA in as many as 8 (8.2%) additional time periods. This result was associated with the inclusion of graupel in the six-class scheme which likely altered latent heat fluxes and leading the minor variation in storm track.

For comparison to the five-class BMPSs and to investigate larger simulated storm track errors in the six-class schemes, a more in-depth analysis of cases 2 and 4 was completed. Figures 24a-d display 500-hPa relative vorticity and geopotential height for

GMA, WSM6, Lin, and NGod simulations for case 2 on 18UTC 16 March 2007. Similar to the five-class BMPS results (Figs. 15b-c), all six-class simulations generated the surface cyclone 400-800 km too far northward as compared to GMA. Comparisons of five- and six-class WSM and NGod runs reveal nearly identical fields of both geopotential height and relative vorticity. As illustrated by Fig. 15, similar to the five-class BMPS simulations, all six-class simulations did not generate enough positive vorticity ($>21 \times 10^{-5} \text{ s}^{-1}$) over southern Alabama where the GMA surface cyclone generated. Instead all simulations favored cyclogenesis near a wide region of 500-hPa latent heat release near the Great Lakes and close to the right-entrance region of the 300-hPa jet (not shown). It is worth noting however that WSM6 exhibited lower track error than other BMPS schemes possibly due to the stronger area of positive 500-hPa relative vorticity ($>12 \times 10^{-5} \text{ s}^{-1}$) over the Gulf of Mexico enhancing upward vertical motion over this region.

Figures 24e-h display 500-hPa relative vorticity and geopotential height for GMA, WSM6, Lin, and NGod for case 4 on 06UTC 2 March 2009. As seen in these figures, every six-class BMPS run maintained a cut-off 500-hPa geopotential height minimum over southeast Georgia, whereas GMA shows it merging with a poleward shortwave height trough at a location roughly 600 km further northward than in any WRF model run. Unlike for case 2, neither the five- and six-class WSM nor Goddard BMPS runs exhibit any notable differences in simulated meteorological fields and no appreciable differences in overall storm track. Due to this similarity, cyclone propagation fell to under 5 m s^{-1} , positive vorticity advection became negligible, and each simulated cyclone eventually become vertically stacked.

Next, maximum simulated cyclone intensity accuracy was assessed by comparing each six-class BMPS runs to GMA. Overall, model simulations over-intensified a nor'easter upwards of 14.2 hPa (case 1; NGod) and under-intensified by as much as 9.5 hPa (case 4; Lin), which equates to 1.18 and -1.77 times the standard deviation of 8.01 hPa. For no one case did all six-class BMPS runs accurately predicted SLP, however all cases did not intensify case 4 enough. With the exception of case 1, the other six cases had at least two runs accurate simulated storm maximum intensity. Between BMPSs the only the Lin and NGod schemes stood out because the former was accurate in 6 out of 8 cases and the latter exhibited an over-intensification tendency. Further analysis of the large-scale environment and SLP deepening rates revealed few notable differences between the various BMPSs run due to their common heritage in Lin. Between BMPSs, the NGod runs was notable because it tended to generate more mid-tropospheric latent heating closer to the cyclone center than the other schemes, which sometimes led it simulated nor'easters to be stronger.

Figures 25 and 26 depict how the six microphysical species (same as the five-class BMPSs and adding graupel) varied among the six-class BMPSs simulations. Figure 25 is the domain 4 shows column-integrated hydrometeor species, and Fig. 26 are the composited, time-averaged hydrometeor profiles. As expected from WP10, Fig. 25 indicated latter to no variability in column-integrated rain and cloud water mixing ratios amongst the six-class BMPS schemes. As for the five-class results, this similarity is attributed to each six-class scheme using a nearly identical approaches for handling liquid hydrometeors originally used in the Lin scheme.

In contrast, column-integrated snow in Fig. 25 varies greatly. Overall, Lin simulated the least column-integrated snow, whereas the OGod simulated the most. As in Tao et al. (2011), the dearth of snow in the Lin scheme is due to its low snow hydrometeor size distribution intercept parameter (thus only large snow particles are counted) and its interactions between snow and other hydrometeors via auto-conversion and accretion. While not as significant as in the Lin scheme, smaller column-integrated snow in WSM6 had the same root cause. On the other hand, column-integrated snow in OGod was roughly 1.5 times higher than the NGod scheme, in line with Lang et al. (2011). Excess snow generation in the OGod scheme is a by-product of its high snow distribution intercept parameter and its snow growth bias via other ice microphysical processes (especially riming). Similarly to the column-integrated plots, in the composite snow profiles (Figure 26), Lin is the driest and OGod the wettest. One striking result was the lower pressure level of maximum snow concentration in the Lin scheme as compared to the other three six-class schemes. This result is attributed the difference in the Lin snow and graupel terminal velocity parameterizations which have been noted to promote the growth of graupel at the expense of snow. As a result, the mid-level and lower levels are devoid of snow because Lin has converted it to graupel.

In Fig. 25, column-integrated graupel in NGod is almost non-existent, whereas the other three schemes generate it in larger quantities. Consistent with WP10, the Lin-based runs generated the most graupel due to its snow and graupel terminal velocity parameterizations favoring graupel growth at the expense of snow. Excess graupel generation in Lin was later addressed when WSM6 was developed (WP10). For WSM6 both the snow and graupel terminal velocities were modified specifically curtail the snow

to graupel conversion rates and increase snow production. In contrast, graupel was nearly non-existent in the NGod scheme due to its lower graupel conversion rates, and is commensurate with Lang et al. (2011). Composite graupel profiles shown in Fig. 26 correspond well to these findings as NGod has the smallest graupel content and Lin the most. Unlike for snow, the level of maximum graupel concentration is nearly identical in all six-class BMPSs.

Similar to the five-class BMPS plots of column-integrated cloud ice, there is considerable variation among the six-class schemes in Fig. 25 where it was highest in WSM6 and lowest in Lin. High levels of cloud ice in WSM6, as in WSM5, as explained previously and in WP10, are attributed excess cloud glaciation in the lower troposphere at temperatures between 0°C and -20°C . Both the Lin and OGod schemes utilized fixed cloud ice size distribution functions, which, in tandem with their ice-conversion processes, produce low column-integrated cloud ice. Comparing these two schemes, the lower cloud ice concentrations Lin can be partly attributed to over conversion of graupel from cloud ice. Compared to the OGod, NGod simulations had lower cloud ice concentrations due to its temperature dependent cloud-ice size distribution functions. Overall, composite, time-averaged cloud ice profiles in Fig. 26 reflect the patterns seen for column-integrated cloud ice. Due to NGod's temperature dependent cloud-ice distribution functions, its level of maximum cloud ice concentration occurs at a lower pressure level. In contrast, other six-class BMPSs have fixed cloud-ice distribution functions which led to each to achieve maximum cloud ice concentration at nearly the same pressure level.

Local-storm environment simulation accuracy was evaluated via energy norm differences (WSM6 - model) for both GMA- (thin lines) and WRF-centered (thick lines) frameworks (Fig. 27). In Fig. 27, positive energy norm differences denote improvement versus WSM6. The bottom of Table 9 displays the number of six-hour periods where the Lin, OGod and NGod runs had a lower energy norm than WSM6 in each framework.

In comparison to WSM6, the Lin, OGod and NGod runs had less simulation error in 26 (42.6%), 20 (32.8%), and 36 (59.0%) out of 61 WRF-centered periods, respectively. At GMA-centered locations these three BMPS had smaller model error than WSM6 in 36 (59.0%), 18 (29.5%), and 45 (73.8%) out of 61 periods. For individual models, WRF-centered energy norm differences were positive in more than 50% of the periods in 3, 1, and 5 of 8 cases for Lin, OGod, and NGod, respectively. Comparing the two Goddard BMPSs revealed that the NGod scheme frequently had less error (45 out of 61 time periods) in either framework. For Lin, NGod simulations had less error in 38 (62.3%) out of 61 time periods. Therefore NGod simulations were the most accurate unlike WP10, who noted WSM6 to be the most accurate. Given their common heritage in Lin, the temperature dependent snow intercept and graupel size distribution in the NGod scheme most likely explains its higher accuracy. This result however may merit some caution given that the small energy norm differences in Fig. 27 and that some cases only had five time periods compared.

Time-averaged and horizontally-integrated energy norms at each vertical level were then calculated in both the WRF- and GMA-centered frameworks (Fig. 28) to detect key sources or error. As seen in Fig. 28, six-class BMPS simulations exhibited three distinct energy norm spikes (~900 hPa, ~750 hPa, and ~200 hPa) which further analysis

showed had the same causes as noted for the five-class BMPS runs previously. In comparison to five-class BMPSs (Fig. 19), six-class BMPSs time-averaged, horizontally-integrated energy norms showed the same overall vertical distribution, and varied only slightly due to differences in simulated storm track. Integrating over the entirety of all four model domains (not shown) a fourth energy norm spike around 300 hPa in association with the upper tropospheric jet stream. Similar to the five-class results, the impact of the upper-level decreased within the smaller inner domains, whereas error with the nor'easter associated, mid-level jet became more important for the inner most domains.

To show how six-class BMPSs impacted WRF simulations as a whole, energy norm were integrated over the entirety of domains 1, 2, 3 and 4. To illustrate how this quantity varied between domains, Fig. 29 depicts energy norm differences for cases 3 (Figs. 29a, 18c, 18e, 18g) and 8 (Figs. 29b, 18d, 18f, 18h). As seen in Fig. 29, OGod and NGod energy norm difference shows nearly identical patterns and of similar magnitude to their five-class BMPS counterparts (Fig. 20). Therefore, the inclusion of graupel did not lead to any significant changes in simulations accuracy. As seen in Fig. 29, NGod simulations fared well against the WSM6, Lin, and OGod simulations and were typically of higher accuracy. OGod simulation appear on-par or slightly worse than both WSM6 and the Lin schemes. Despite the inclusion of graupel, the OGod scheme still exhibited greater error on coarser domains. Meanwhile Lin-based simulations appear on average to be slightly better or on-par with WSM6.

As compared to WSM6, Lin, OGod and NGod simulations had smaller energy norms in at most 121 (48.8%; NGod), 138 (55.6%; NGod), 151 (60.9%; Lin), and 133

(53.6%; Lin) out of 248 six-hour periods for domains 1, 2, 3 and 4, respectively (Table 9). Lin simulations demonstrated lower error than WSM6 in 2, 3, 7, and 3 out of 8 cases, on domains 1-4, respectively and in as many as 151 (60.9%) out of 248 periods. Next, on domains 1-4, OGod outperformed WSM6 in 1, 0, 1, and 4 out of 8 cases, respectively and in as many as 112 (45.2%) out of 248 periods. NGod simulations exhibited lower energy norms than WSM6 in 4, 4, 3, and 5 out of 8 cases, respectively and in as many as 138 (55.6%) out of 248 periods. Directly comparing OGod and NGod revealed strikingly similar results to the five-class BMPS results with up to 83.9% (domain 2) of time periods and all but case 4 having lower error in the latter. These results show NGod simulations were marginally better than WSM6 on all model domains, except the 45 km resolution domain 1. These results are unlike WP10 which suggested WSM6 to be marginally superior to all other six-class BMPSs on their 25 and 5 km resolution WRF domains. As compared to OGod, NGod outperformed it on all but domain 4 where it improved 127 (51.2%) out of 248 periods relative to NGod. Lin was marginally better than NGod on domain 3 (126 out of 248 periods) and was equal to it on domain 4. Given that energy norm differences, with only a few exceptions (domain 3, case 8), were often 2 orders of magnitude less than the energy norm itself suggest only small changes to overall simulation accuracy. Given all the above, the energy norm results suggest that NGod simulations produced the best overall simulations given both its consistency from domain to domain and its model error being either roughly equivalent to or marginally better than all other BMPSs.

For comparison to the energy norm results and to determine which meteorological fields were most sensitive to WRF model version, RMSE calculations were completed

the same five RMSEs as noted previously and are summarized in Table 9. Values shown in Table 7 denote the number of periods where each six-class BMPS had a lower RMSE than WSM6. Similar to the five-class BMPSSs, RMSE results were not always consistent with the energy norm results, especially for 300-hPa winds and 500-hPa geopotential height. To illustrate these results graphically, Fig. 30 displays the energy norm (Fig. 30a) and five RMSEs (Figs. 30b-e) from case 6 from domain 2. This case was selected both because of its representativeness, but also to afford comparison to the five-class BMPSSs (Fig. 21). Despite the presence of graupel, the six-class Goddard RMSE differences are of similar shape and magnitude to their five-class counterparts suggesting its impact upon simulations to be minor. As for the Lin scheme, its RMSE differences in Fig. 30 were fairly consistent with its energy norm results. As suggested in Fig. 30 however, the magnitude of the RMSE differences relative to the meteorological parameters they represent were often 2 orders of magnitude smaller. Although, it should be noted that domain 4 300-hPa wind RMSEs were sometimes within 1 order of magnitude of the average, 300-hPa wind speed.

To summarize the six-class scheme results using the most comprehensive metric, the energy norm, NGod simulation accuracy was the most consistent (6.7% variability) between model domains and was the most accurate on domain 2. For domain 1, WSM6-based simulations were marginally better than NGod (6 time periods), notably better than Lin (44 more time periods), and far better than Lin (170 more periods). On domains 3 and 4, the Lin scheme showed marginally outperformed all other BMPSSs with the exception of where it tied NGod on domain 4. Combining all 24 tests (energy norm and RMSE from all domains), show Lin simulations had the highest accuracy versus WSM6 (11 out of 24

tests). However, 7 out of 11 victories originated from only 500-hPa geopotential height and 850-hPa temperature RMSEs which may suggest better handling of mid-tropospheric latent heat release and temperature advection. Finally, with respect to the local storm environment, NGod simulations outperformed all other BMPS in 59%, 62.3%, and 73.8%, of all periods relative to WSM6, Lin, and OGod, respectfully. The above results highlight that temperature dependent snow, cloud ice and graupel function indeed lead to better simulations of the dynamics that underpin nor'easters, however as seen by the domain 3 results, this additional model complexity does not necessarily lead to better simulations under all conditions.

3.4. Conclusions

The impact of five- and six-class bulk microphysics schemes (BMPSs) on WRF 3.3 simulations of eight nor'easter cases was investigated. The five-class schemes included WSM5, OGod-2ice, and NGod-2ice, and the six-class schemes were WSM6, Lin, OGod-3ice, and NGod-3ice. Model simulations were conducted for 180 hours, starting roughly 72 hours prior to the first precipitation impacts in the highly populated Mid-Atlantic US and associated cyclogenesis had boundary conditions derived from GFS model forecasts. Simulation accuracy (except for simulated hydrometeor mixing ratios) was assessed by comparing each simulation to GMA.

Analysis of SLP-based storm track revealed BMPS choice generally had little impact on storm track (exception WSM schemes, cases 2 and 8), intensity, propagation speed, or key meteorological fields (i.e., 500-hPa geopotential height, relative vorticity). Track errors relative to GMA were typically between 240-250 km and 255-305 km for

five-class and six-class schemes respectively. The inclusion of graupel to the WSM and Goddard schemes only introduced up to a 28 km differences in average track error. Consistent with WP10 and Tao et al. (2011), inter-WRF track variability for six-class BMPSs was small (50km). Most WRF simulations overall exhibited a leftward bias (>70.8% of simulations) and lagged GMA (>58.3% of simulations). For individual BMPSs, all five and six-class schemes with the exception of WSM6 favored a leftward bias or lagged GMA at least 53.3% and as much as 66.7% of all time periods. This exception likely stems from altered latent heat fluxes in cases 2 and 8 where large-scale forcing was weak.

Between five-class BMPS simulations, nor'easters were up to 15.6 hPa too strong and 9.5 hPa too weak relative to GMA. Overall, all BMPSs accurately predicted SLP in cases 6 and 7, over-intensified case 1, and under-intensified case 3. For the other four cases, cyclone intensity was accurate in 2 out of 3 WRF simulations for each case. For six-class BMPSs, simulated nor'easters were up to 14.2 hPa too strong and 9.5 hPa too weak. Unlike the five-class BMPS, in not one case were all four, six-class BMPSs accurate, but they all produced too weak a nor'easter in case 4. Despite these results, no BMPS scheme (five or six class) showed significant variations in simulated meteorological fields (i.e., SLP, 500-hPa latent heating, 500-hPa geopotential heights) which is due to their common heritage in Lin.

Column-integrated and time-integrated profiles were used to investigate hydrometeor concentration variability for both five and six-class BMPSs. Because all BMPSs used Lin as the basis for simulating rain and cloud water concentrations, no notable variability existed between BMPS schemes. On the other hand, each BMPS

however did make adjustments to how each handles frozen hydrometeors. For snow, the OGod schemes generated the most snow, whereas the Lin scheme generated the least. Cloud ice concentrations demonstrated considerable inter-BMPS variability. Specifically, WSM6 tended to over glaciate cloud in the lower troposphere which resulted in excessive cloud ice concentrations, whereas for Lin and NGod these values were much reduced. Finally, graupel concentrations ranged from almost non-existent in NGod to over-production in Lin simulations.

The energy norm was a useful metric for evaluating the performance of the five- and six-class BMPS schemes in both the local storm environment and for the entire WRF simulation. For the local-storm environment, results suggest NGod simulations were the most accurate five-class BMPS. It exhibited less error than WSM5 and OGod in 34 (57.6%) and 41 (69.5%) out of 59 time periods respectively. For six-class BMPSs. Similarly, NGod produced the most accurate simulations of the local-storm environment out of all four six-class BMPSs. Energy norms from NGod simulations were lower WSM6, Lin, and OGod in 36 (59.0%), 45 (73.7%), and 38 (62.3%) out of 61 periods, respectfully. Given that each BMPS has a heritage in the Lin scheme the success of the NGod scheme can only be attributed to its temperature dependent snow, cloud ice, and graupel functions which existed in no other scheme.

For all BMPSs there were four distinct energy norm spikes at ~900 hPa, ~750 hPa, ~300 hPa, and ~200 hPa. These spikes were associated with errors in the planetary boundary layer height and the low-level jet, mid-tropospheric jet intensity and positioning, upper-level jet stream intensity and positioning, and tropopause height, respectively. The largest overall contributor was 300-hPa winds on coarser domains and

750 hPa winds on the inner domains due to the location, magnitude, and spatial scales of both jets relative to each model domain.

Evaluation of energy norm and RMSE over the entirety of all four domains demonstrated that while NGod simulations handled the local-storm environment well, the same was not necessarily true for the entire simulation. Specifically, in both five- and six-class simulations, the NGod scheme never had the lowest RMSE for either 300-hPa winds or for 500-hPa geopotential height on any model domain. Given the levels involved, increased model error in NGod are possibly associated with its temperature dependent snow, cloud ice, and graupel concentrations and impacted latent heating at these levels. This claim is further substantiated by OGod being more accurate than NGod in more than 58% of all time periods for all four model domains. Overall, NGod energy norm differences relative to WSM6 were the most consistent of all BMPS (6.7% time period change between domains). NGod simulations exhibited lower energy norms in more than 50% of all time periods for domains 2-4 relative to WSM6, domain 1-2 relative to Lin, and domains 1-3 relative to OGod. Despite these results, in instances where NGod was best overall or not, both NGod energy norm and RMSE differences were only marginally better or worse than other BMPSs.

Overall, this study has shown that BMPS choice typically led minor differences in overall WRF simulation accuracy with the exception of the resolution dependent OGod scheme. When focusing exclusively upon a complex dynamical situation that is a nor'easter, however, various assumptions made by each BMPS led to notable differences in the distribution of hydrometeor species and likely precipitation patterns as well. Given the lack of notable large-scale simulation differences, the unique temperature dependent

cloud ice, snow, and graupel concentrations made by the NGod scheme likely led to its success in simulation the nor'easter environment. These assumptions however do not always lead to better WRF simulations of the entire environment. Most notably, the Lin scheme makes none of these temperature-dependent assumptions and is the oldest BMPS used yet its accuracy was comparable to NGod. Therefore using more complex BMPSs in WRF does not guarantee better model simulation under any circumstance, yet in complex weather situations choosing an appropriate BMPS may prove instrumental to accurate weather prediction.

3.5. Future work

This work broadly addressed how five- and six- class BMPSs impacted WRF simulations of nor'easters, but more work could still be completed in follow-up studies. Future studies could investigate double-moment BMPSs and whether having a variable number concentration would lead to notable changes in simulation accuracy. Additionally, this study did not investigate how BMPS choice impacted many key aspects of the nor'easter (such as precipitation patterns). Another potential research route would investigate how BMPS choice impacts other weather events such as tropical cyclones. Finally, given the challenges in verifying high-resolution forecasts against lower resolution ground truth data, additional verification methods such as object-oriented (Marzban and Sandgathe 2006) or fuzzy verification (Elbert 2008) could be utilized.

Table 4: Microphysics schemes used in the present study and the included microphysics mixing ratios denoted with an “X”.

Microphysics	QV	QC	QI	QR	QG	QS	Citation
WSM5	X	X	X	X		X	Hong et al. (2004)
WSM6	X	X	X	X	X	X	Hong and Lim (2006)
Lin	X	X	X	X	X	X	Lin et al. (1983); Rutledge and Hobbs (1984)
OGod	X	X	X	X	X*	X	Lang et al. (2007)
NGod	X	X	X	X	X*	X	Lang et al. (2011)

*Has options for both five- and six-class configuration

Table 5: Nor'easter case list. The NESIS number is included for storm severity reference. The last two columns denote the first and last times for each model run.

Case Number	NESIS	Event Dates	Model Run Start Date	Model Run End Date
1	N/A	22-24 Nov. 2006	11/19 12UTC	11/27 00UTC
2	2.54	15-17 Mar. 2007	3/12 18UTC	3/20 06UTC
3	N/A	15-17 Apr. 2007	4/12 06UTC	4/19 18UTC
4	1.65	1-2 Mar. 2009	2/26 12UTC	3/6 00UTC
5	N/A	15-16 Oct 2009	10/12 12UTC	10/20 00UTC
6	4.03	19-20 Dec. 2009	12/16 06UTC	12/23 18UTC
7	4.38	4-7 Feb. 2010	2/2 18UTC	2/10 06UTC
8	N/A	12-14 Mar. 2010	3/9 06UTC	3/16 18UTC

Table 6: GMA-relative storm track bias analysis for five-class microphysics schemes. Values indicate the total the number six-hour time periods from all cases where each simulation exhibited the shown track bias relative to GMA at each time period.

	Left		Right		Key
Lead	21	22	22	18	
	20		23		
Lag	33	36	16	19	
	32		22		

WSM5 Ngod Ogod

Table 7: Values denote the number and percentage of 6-hour periods where the energy norm or RMSE was lower than WSM5 for each simulation. The last section indicates the number of periods where BMPS-based energy norms were lower than WSM5 within a 600 km box centered on GMA and each simulated storm, respectively. Bolded values indicate the best WRF simulation (as compared to WSM5) for each analysis component.

Domain 1 Performance Analysis vs WSM5 (248 Total Times)	Ogodd-2ice	Ngodd-2ice
Energy Norm	38 (15.3%)	127 (51.2%)
RMSE 2 PVU Pot T	110 (44.4%)	127 (51.2%)
RMSE 300-hPa Winds	137 (55.2%)	94 (37.9%)
RMSE 500-hPa Geo Hght	130 (52.4%)	96 (38.7%)
RMSE 850-hPa Temps	85 (34.3%)	134 (54.0%)
RMSE SLP	136 (54.8%)	161 (64.9%)
Domain 2 Performance Analysis vs WSM5 (248 Total Times)	Ogodd-2ice	Ngodd-2ice
Energy Norm	15 (6.0%)	137 (55.2%)
RMSE 2 PVU Pot T	79 (31.9%)	138 (55.6%)
RMSE 300-hPa Winds	147 (59.3%)	100 (40.3%)
RMSE 500-hPa Geo Hght	118 (47.6%)	79 (31.9%)
RMSE 850-hPa Temps	92 (37.1%)	96 (38.7%)
RMSE SLP	139 (56.0%)	130 (52.4%)
Domain 3 Performance Analysis vs WSM5 (248 Total Times)	Ogodd-2ice	Ngodd-2ice
Energy Norm	90 (36.3%)	135 (54.4%)
RMSE 2 PVU Pot T	66 (26.6%)	123 (49.6%)
RMSE 300-hPa Winds	129 (52.0%)	119 (48.0%)
RMSE 500-hPa Geo Hght	122 (49.2%)	97 (39.1%)
RMSE 850-hPa Temps	130 (52.4%)	89 (35.9%)
RMSE SLP	129 (52.0%)	118 (47.6%)
Domain 4 Performance Analysis vs WSM5 (248 Total Times)	Ogodd-2ice	Ngodd-2ice
Energy Norm	118 (47.6%)	122 (49.2%)
RMSE 2 PVU Pot T	85 (34.3%)	115 (46.4%)
RMSE 300-hPa Winds	126 (50.8%)	109 (44.0%)
RMSE 500-hPa Geo Hght	127 (51.2%)	110 (44.4%)
RMSE 850-hPa Temps	125 (50.4%)	100 (40.3%)
RMSE SLP	121 (48.8%)	113 (45.6%)
Domain 2 Storm Performance Analysis vs WSM5 (59 Total Times)	Ogodd-2ice	Ngodd-2ice
Energy Norm GFS Rel	20 (33.9%)	42 (71.2%)
Energy Norm WRF Rel	26 (44.1%)	34 (57.6%)

Table 8: Same as Table 6, except for six-class microphysics schemes.

	Left		Right			
Lead	21	22	22	18	Key	
	20		23		WSM5	Ogod
Lag	33	36	16	19	Ngod	
	32		22			

Table 9: Values denote the number and percentage of 6-hour periods where the energy norm or RMSE was lower than WSM6 for each simulation. The last section indicates the number of periods where BMPS-based energy norms were lower than WSM6 within a 600 km box centered on GMA and each simulated storm, respectively. Bolded values indicate the best WRF simulation (as compared to WSM6) for each analysis component.

Domain 1 Performance Analysis vs WSM6 (248 Total Times)	Lin	OGod	NGod
Energy Norm	102 (41.1%)	39 (15.7%)	121 (48.8%)
RMSE 2 PVU Pot T	114 (46.0%)	113 (45.6%)	122 (49.2%)
RMSE 300-hPa Winds	137 (55.2%)	146 (58.9%)	90 (36.3%)
RMSE 500-hPa Geo Hght	196 (79.0%)	141 (56.9%)	102 (41.1%)
RMSE 850-hPa Temps	127 (51.2%)	86 (34.7%)	129 (52.0%)
RMSE SLP	159 (64.1%)	139 (56.0%)	168 (67.7%)
Domain 2 Performance Analysis vs WSM6 (248 Total Times)	Lin	OGod	NGod
Energy Norm	114 (46.0%)	24 (9.7%)	138 (55.6%)
RMSE 2 PVU Pot T	116 (46.8%)	84 (46.8%)	134 (54.0%)
RMSE 300-hPa Winds	123 (49.6%)	151 (60.9%)	104 (41.9%)
RMSE 500-hPa Geo Hght	163 (65.7%)	117 (47.2%)	77 (31.0%)
RMSE 850-hPa Temps	136 (54.8%)	103 (41.5%)	97 (39.1%)
RMSE SLP	139 (56.0%)	142 (57.3%)	131 (52.8%)
Domain 3 Performance Analysis vs WSM6 (248 Total Times)	Lin	OGod	NGod
Energy Norm	151 (60.9%)	85 (34.3%)	129 (52.0%)
RMSE 2 PVU Pot T	109 (44.0%)	82 (33.1%)	114 (46.0%)
RMSE 300-hPa Winds	128 (51.6%)	131 (52.8%)	115 (46.4%)
RMSE 500-hPa Geo Hght	147 (59.3%)	131 (52.8%)	95 (38.3%)
RMSE 850-hPa Temps	170 (68.5%)	139 (56.0%)	87 (35.1%)
RMSE SLP	136 (54.8%)	129 (52.0%)	115 (46.4%)
Domain 4 Performance Analysis vs WSM6 (248 Total Times)	Lin	OGod	NGod
Energy Norm	133 (53.6%)	112 (45.2%)	132 (53.2%)
RMSE 2 PVU Pot T	109 (44.0%)	87 (35.1%)	120 (48.4%)
RMSE 300-hPa Winds	127 (51.2%)	127 (51.2%)	110 (44.4%)
RMSE 500-hPa Geo Hght	128 (51.6%)	124 (50.0%)	101 (40.7%)
RMSE 850-hPa Temps	160 (64.5%)	138 (55.6%)	104 (41.9%)
RMSE SLP	140 (56.5%)	143 (57.7%)	117 (47.2%)
Domain 2 Storm Performance Analysis vs WSM6 (61 Total Times)	Lin	OGod	NGod
Energy Norm GFS Rel	36 (59.0%)	18 (29.5%)	45 (73.8%)
Energy Norm WRF Rel	26 (42.6%)	20 (32.8%)	36 (59.0%)

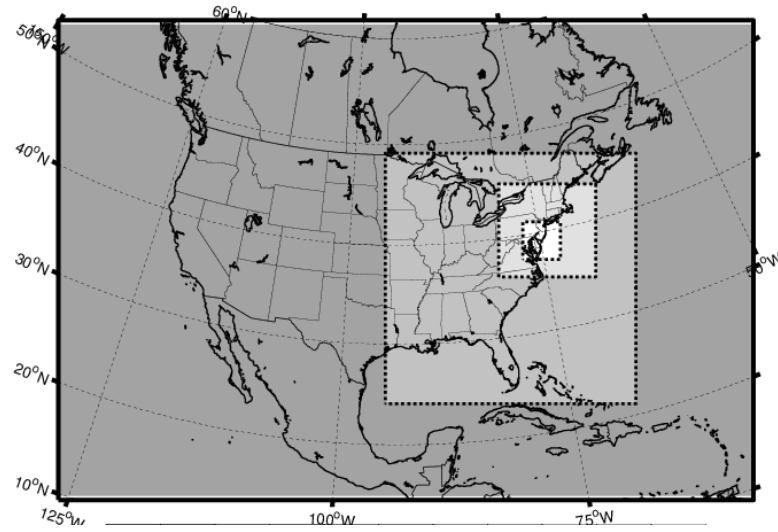


Fig. 12: Nested WRF configuration used in simulations. Horizontal resolution for domains 1, 2, 3, and 4 are 45, 15, 5, and 1.667 km, respectively.

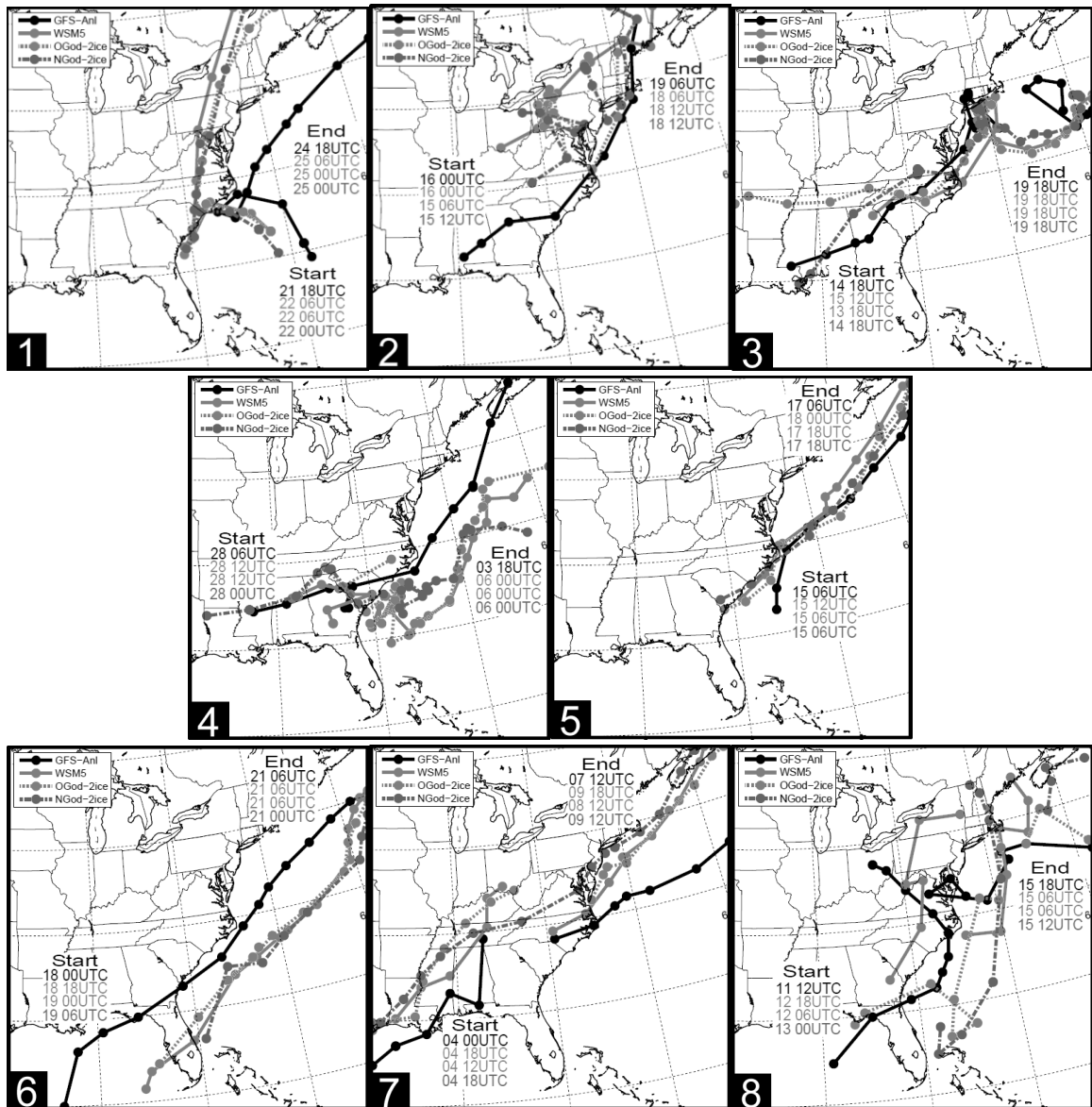


Fig. 13: Storm tracks from GMA and the WRF 6-class microphysics runs. Line legend is shown on the upper-left of each plot. Shown symbols indicate simulated storm position every six hours. White numbers indicate case number.

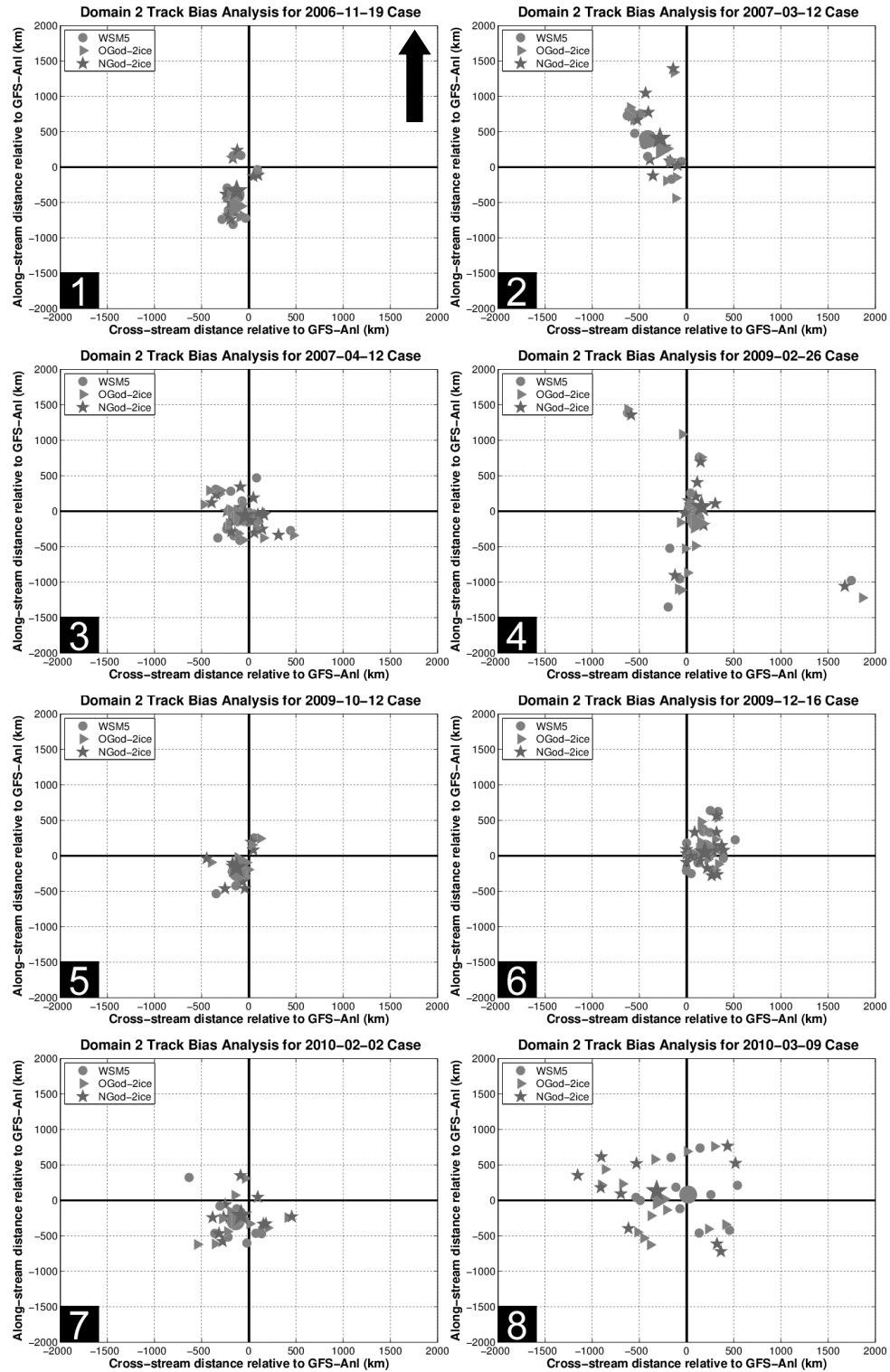


Fig. 14: WRF five-class microphysics run storm position bias as compared to GMA for all eight cases. The black arrow in panel 1 indicates the GMA storm motion direction for all panels. Shown symbols represent WRF position bias every six hours (smaller symbols) and their mean (large symbols). Case number is indicated with a white number.

GMA

WSM5

NGod

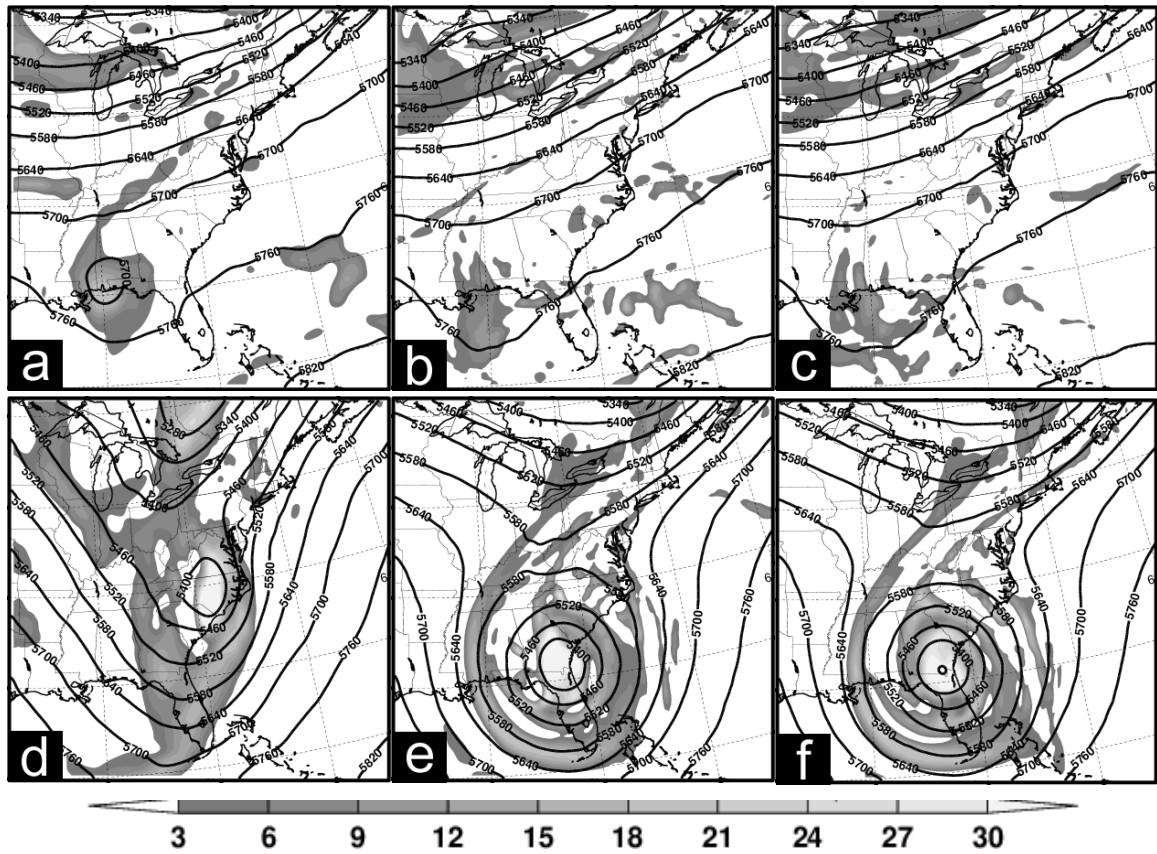


Fig. 15: 500-hPa positive relative vorticity (fills; 10^{-5} s^{-1}) and 500-hPa geopotential height (contours; m) on (a-c) 00 UTC 16 March 2007 and (d-f) 06 UTC 2 March 2009.

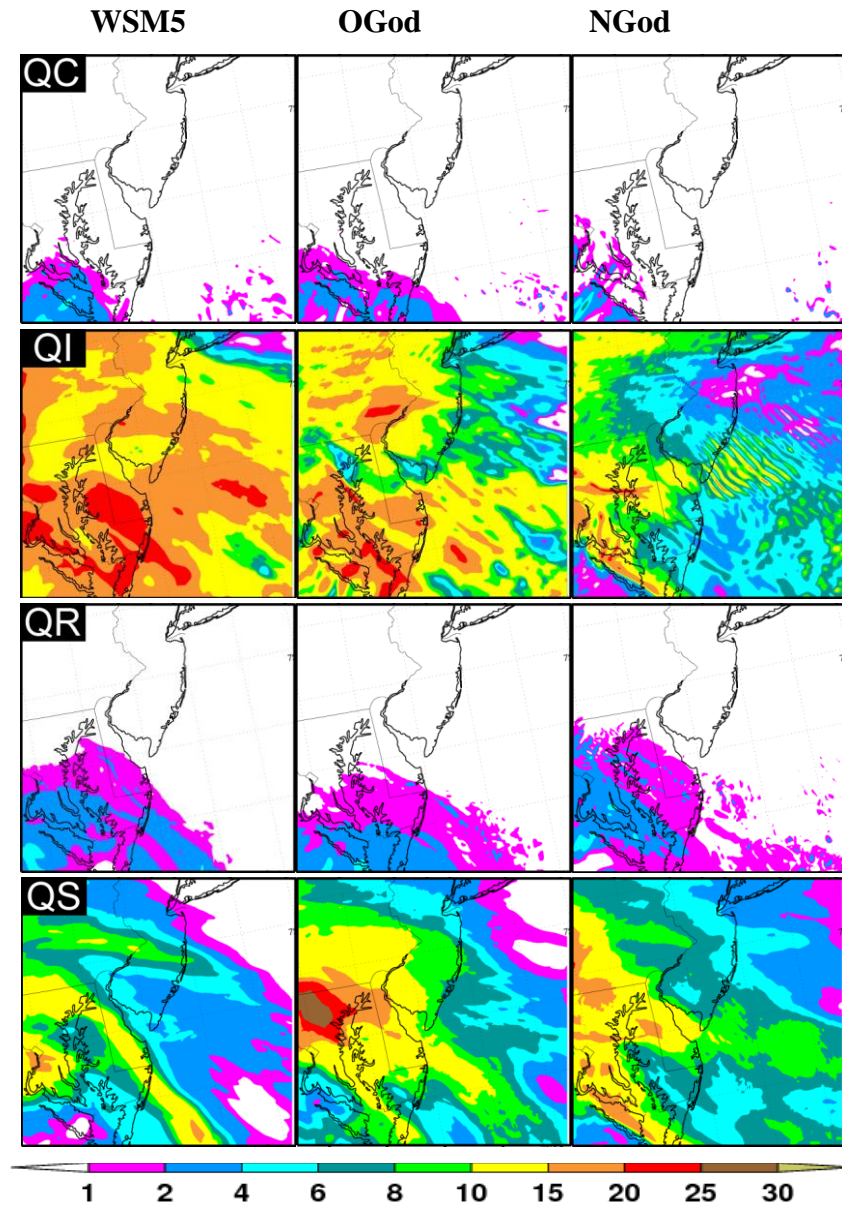


Fig. 16: Domain 4, column-integrated mixing ratios (kg m^{-2}) on 6 Feb 2010 at 06UTC. Shown abbreviations are QC = cloud water mixing ratio, QI = cloud ice mixing ratio, QR = rain mixing ratio, and QS = snow mixing ratio. All shown values are in kg m^{-2} except QI which is scaled by 10^1 .

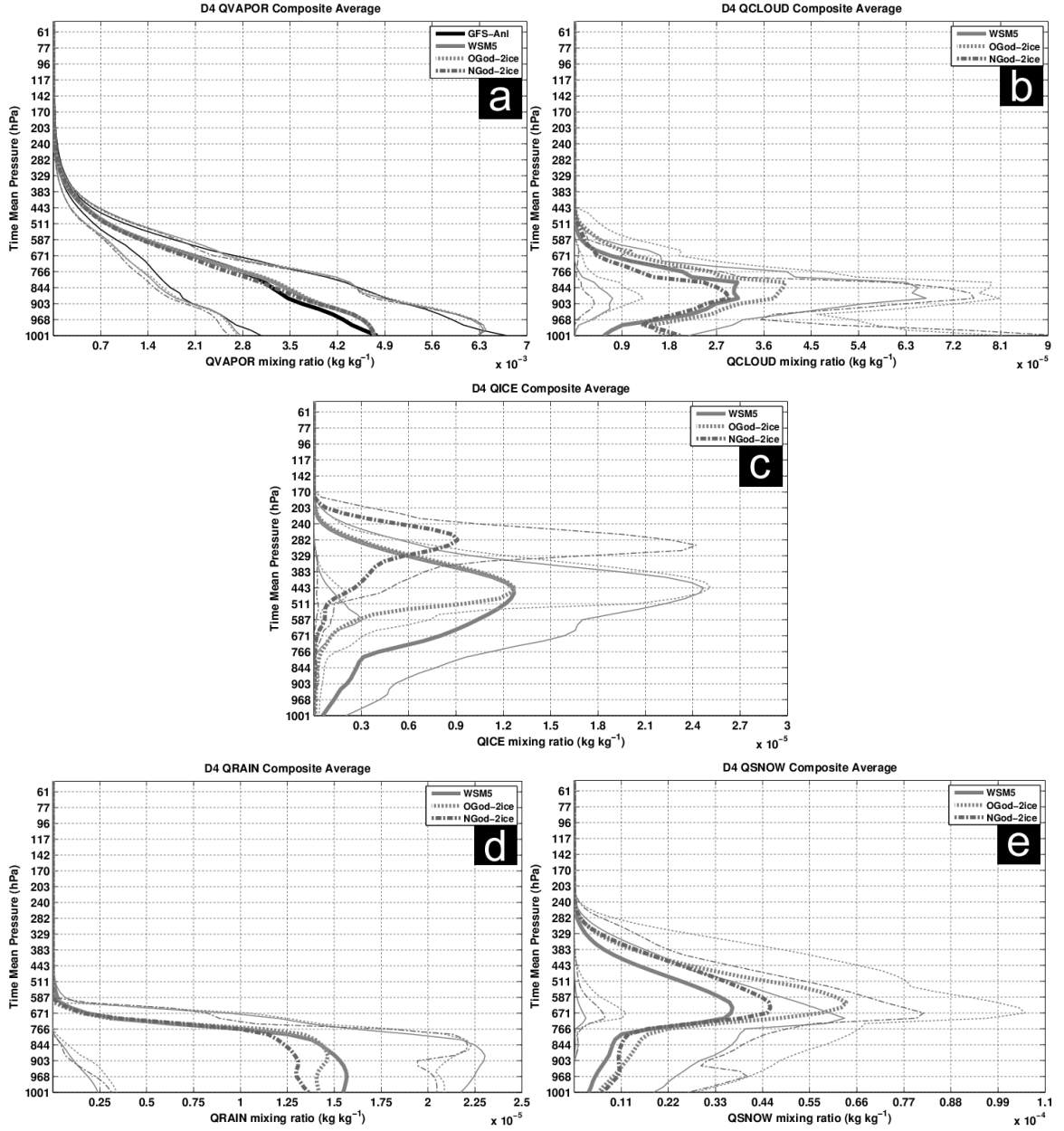


Fig. 17: Domain 4 time-averaged, case composite mixing ratios (kg kg⁻¹) from all 5-class microphysics scheme model runs. The eight-case average mixing ratio and the highest and lowest mean mixing ratio are indicated by thick and thin lines, respectively. Shown mixing ratio composites include those for (a) water vapor, (b) cloud water, (c) cloud ice, (d) rain and (e) snow.

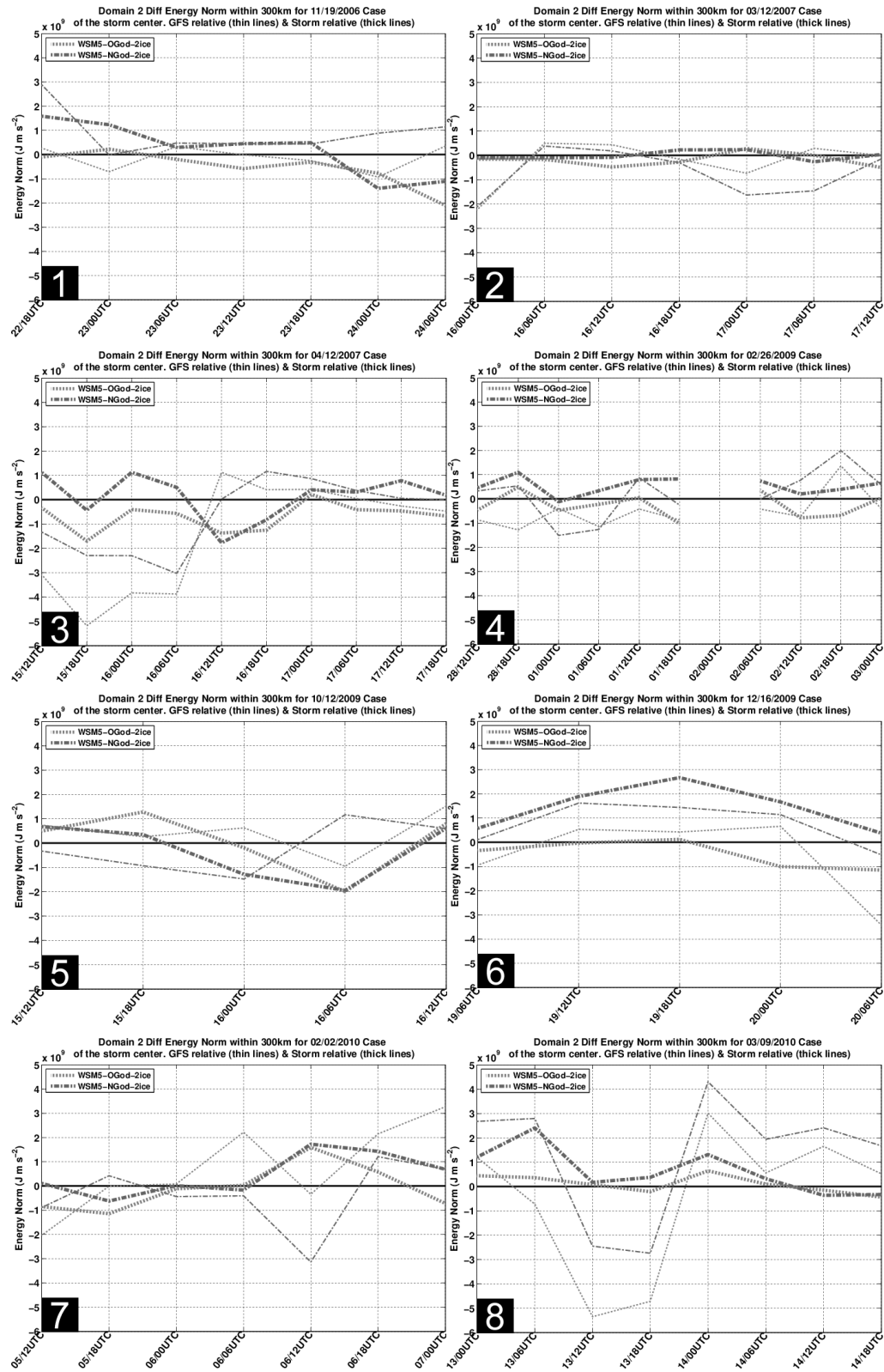


Fig. 18: Domain 2 energy norm differences for all cases within 300 km of the GMA storm center (thin lines) and each model simulated center (thick lines). Shown differences are relative to WSM5 with positive values denoting improvement.

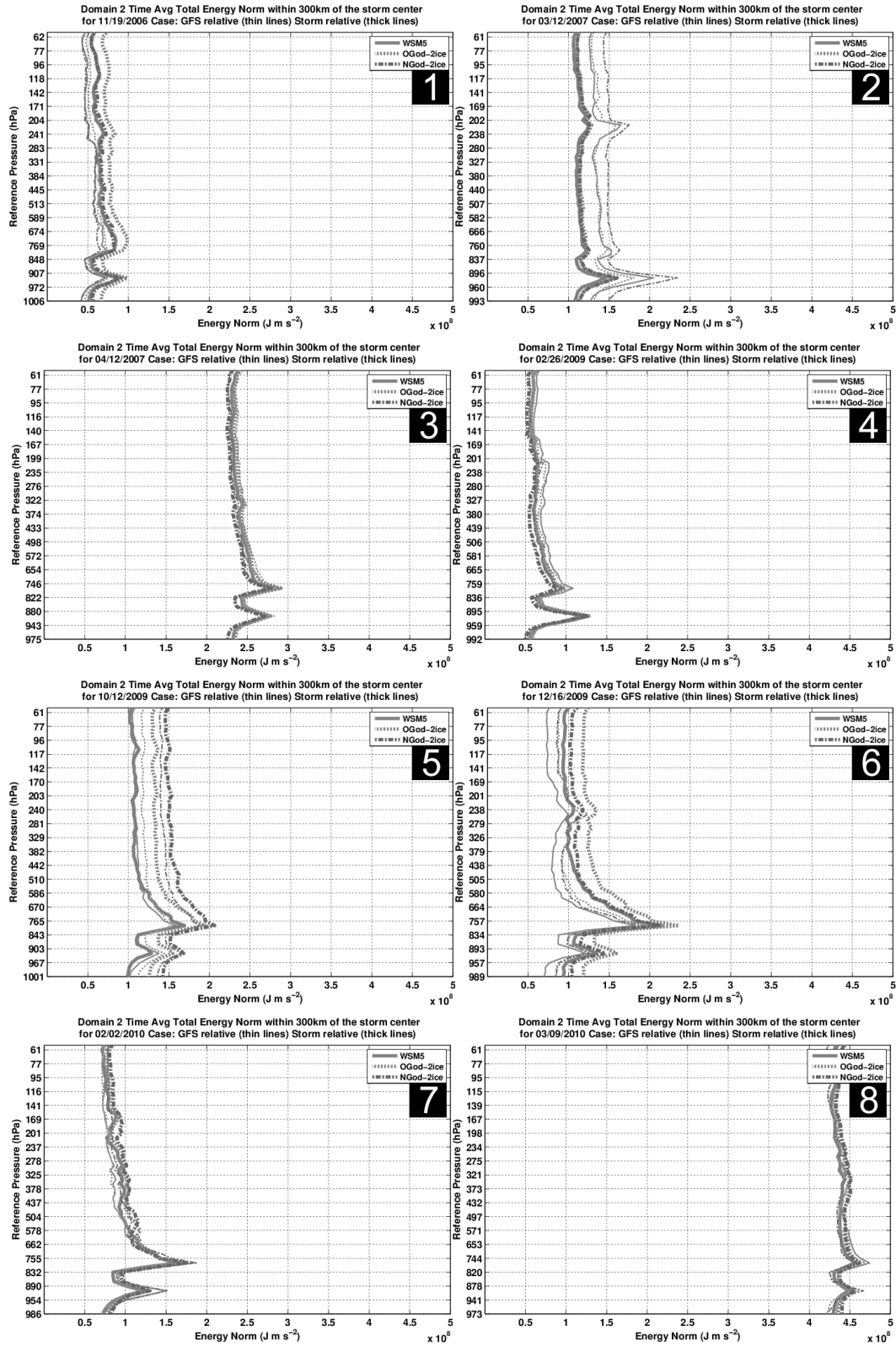


Fig. 19: 180-hr averaged domain 2 energy norm profiles for all cases within 300km of both GMA (thin lines) and each model simulated center (thick lines). The line legend for all 5-Class microphysics schemes is shown in the top right of each panel.

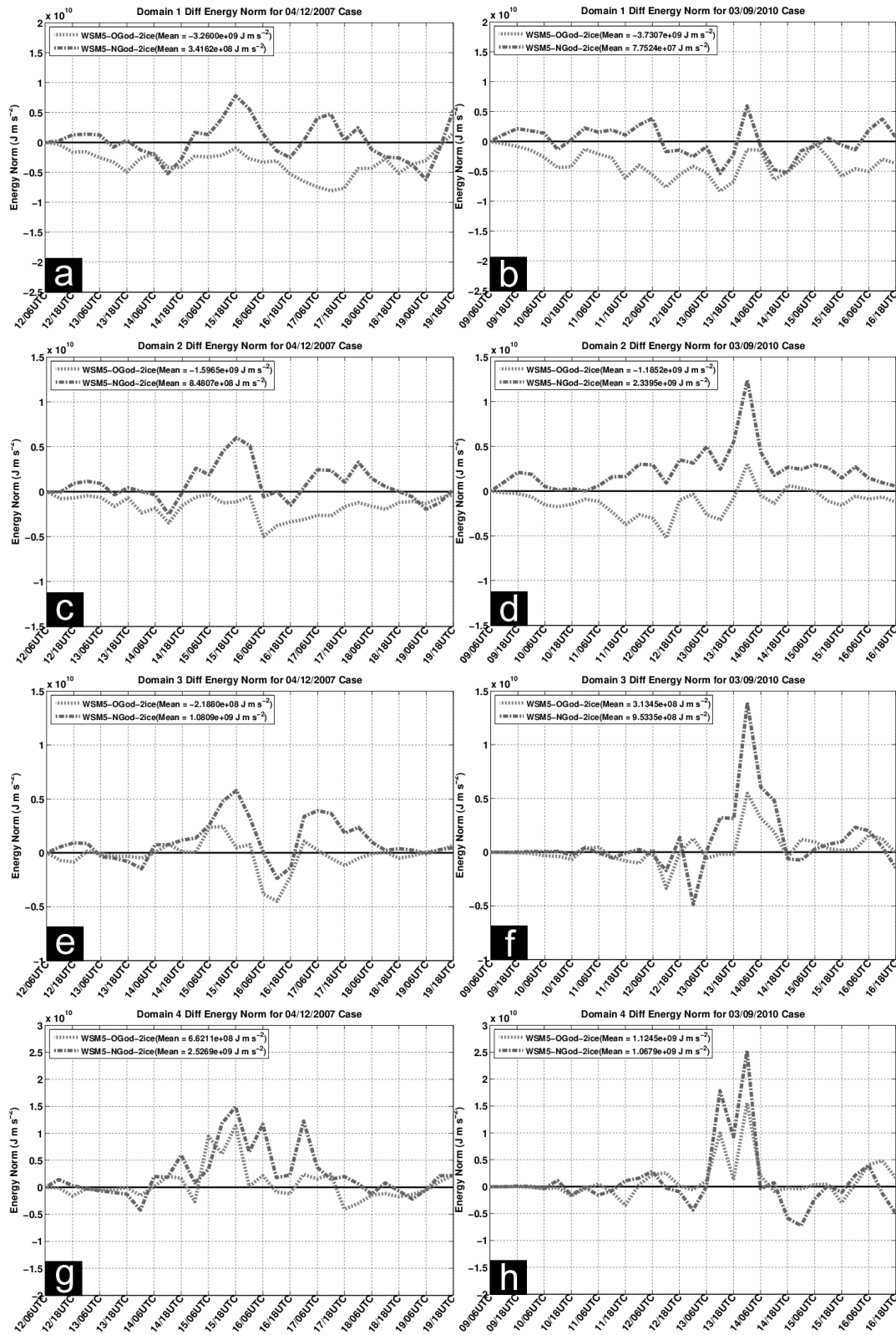


Fig. 20: Energy norm differences within the entirety of domains 1, 2, 3 and 5 for (a, c, e, g) cases 3 and (b, d, e, h) 8. Shown differences are relative to WSM5. Positive values denote improvement over WSM5.

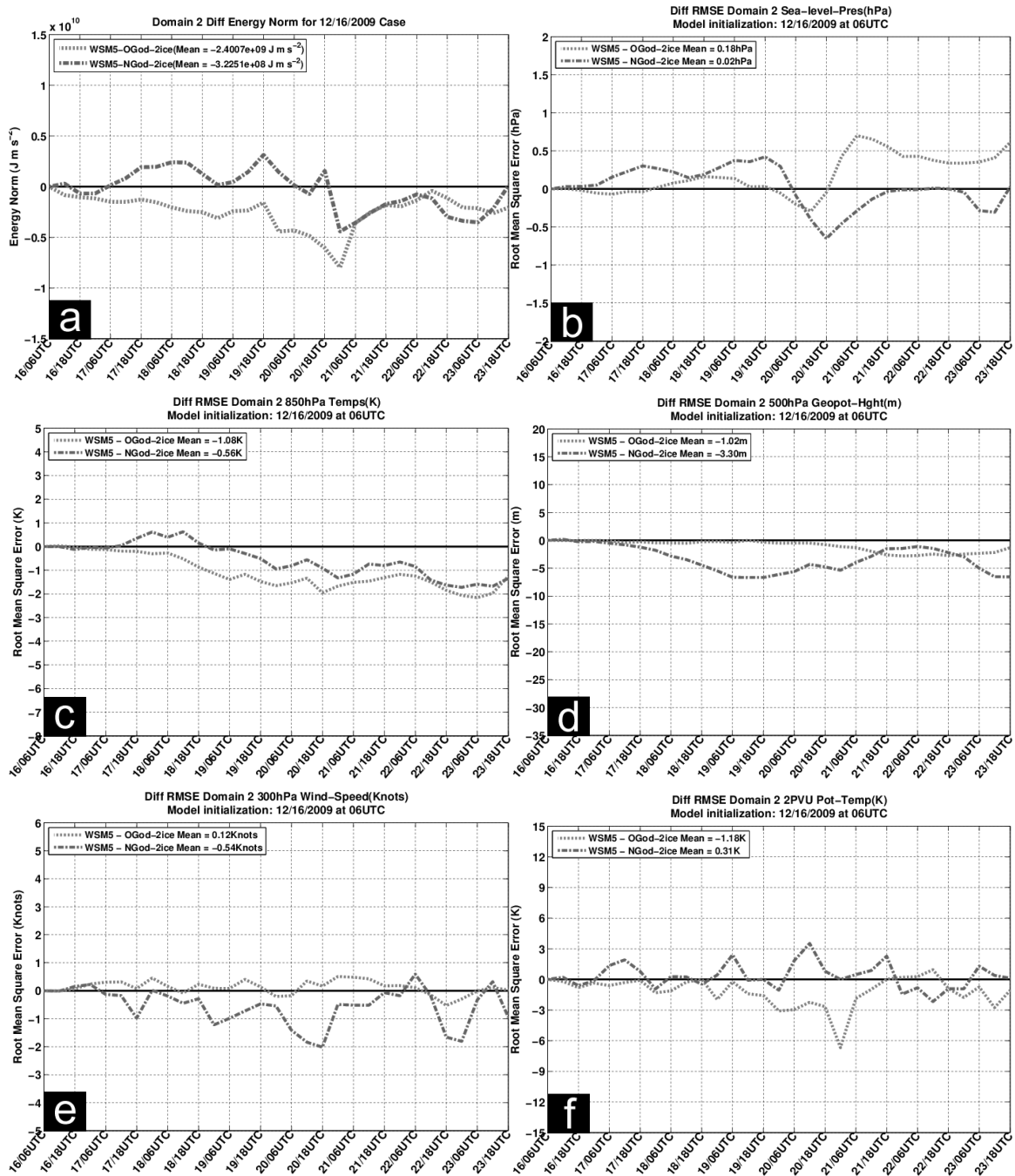


Fig. 21: Difference in domain 2 (a) energy norm, (b) sea-level pressure, (c) 850-hPa temperature, (d) 500-geopotential height, (e) 300-hPa winds, (f) 2-PVU potential temperature for case 6. Shown differences are relative to WSM5. Positive values denote improvement over WSM5.

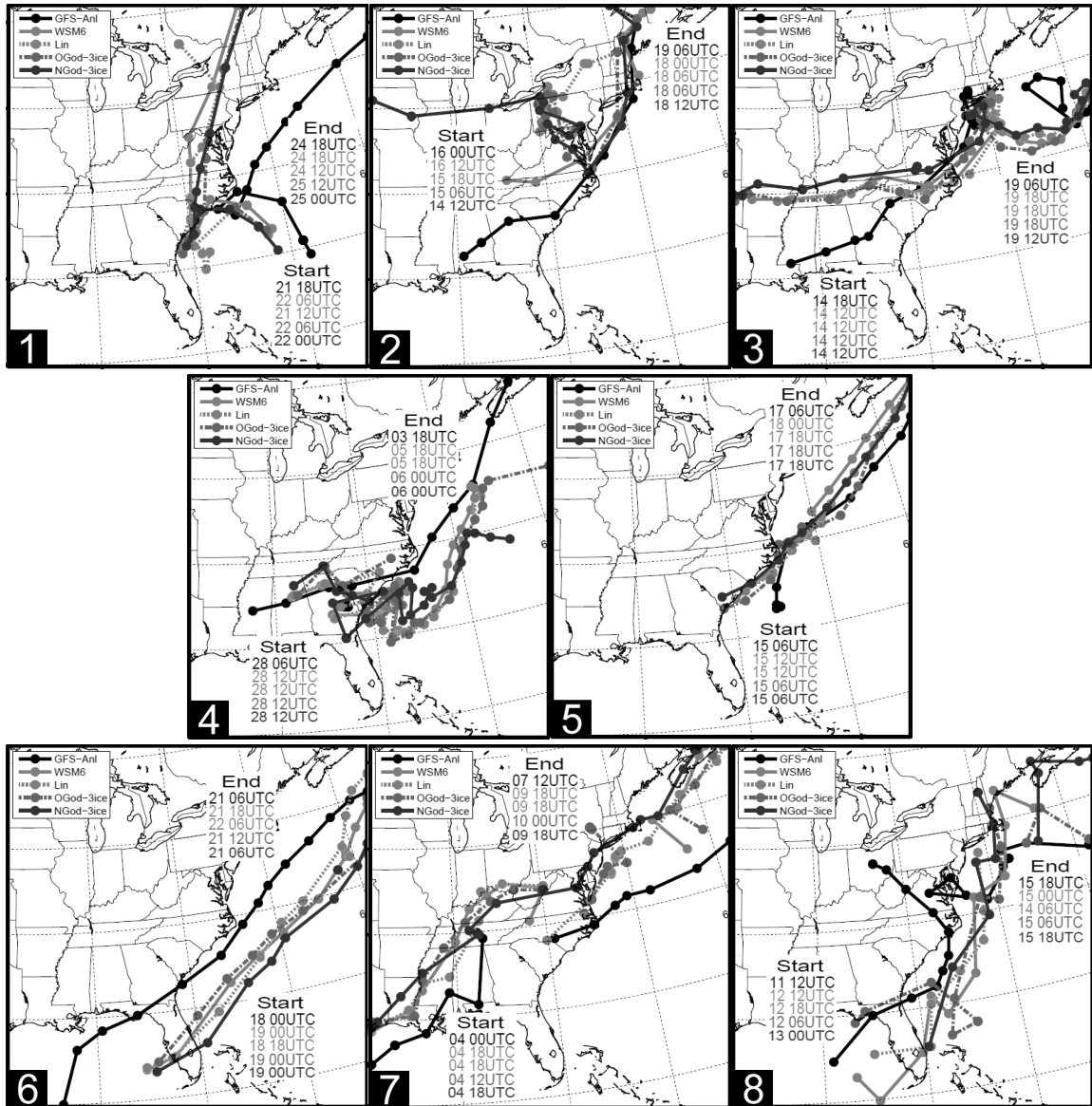


Fig. 22: Sea-level pressure based storm tracks for all eight cases from GMA and each of the six-class BMPs. Line legend is shown on the upper-left of each plot. Storm position is indicated once every six hours with the shown dots.

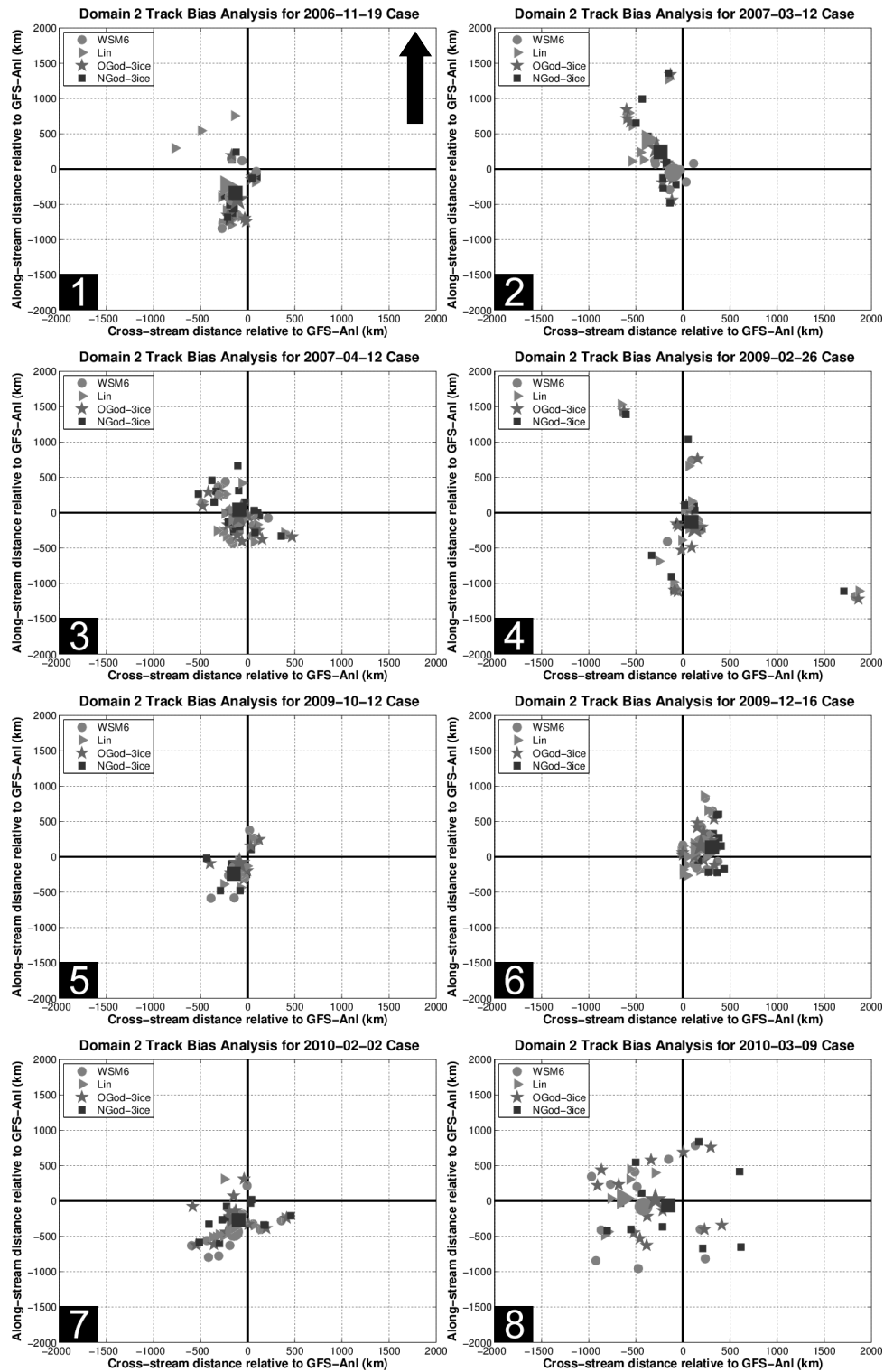


Fig. 23: WRF six-class microphysics run storm position bias as compared to GMA for all eight cases. The black arrow in panel 1 indicates the GMA storm motion direction for all panels. Shown symbols represent WRF position bias every six hours (smaller symbols) and their mean (large symbols). Case number is indicated with a white number.

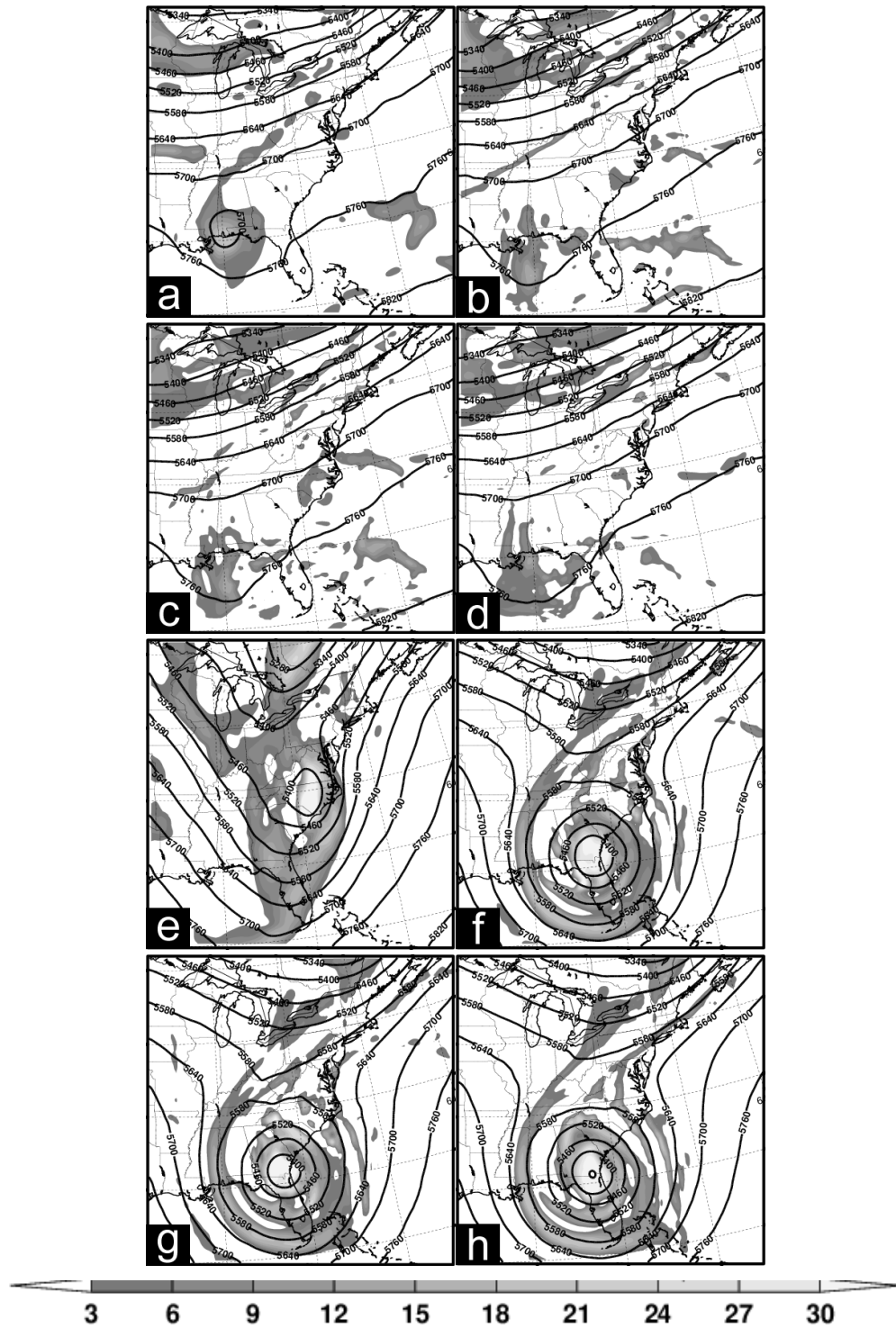


Fig. 24: 500-hPa positive relative vorticity (fills; 10^{-5} s^{-1}) and 500-hPa geopotential height (contours; m) on (a-c) 00 UTC 16 March 2007 and (d-f) 06 UTC 2 March 2009. Shown plots are from (a,e) GMA, (b,f) WSM6, (c,g) Lin, and (d,h) NGod.

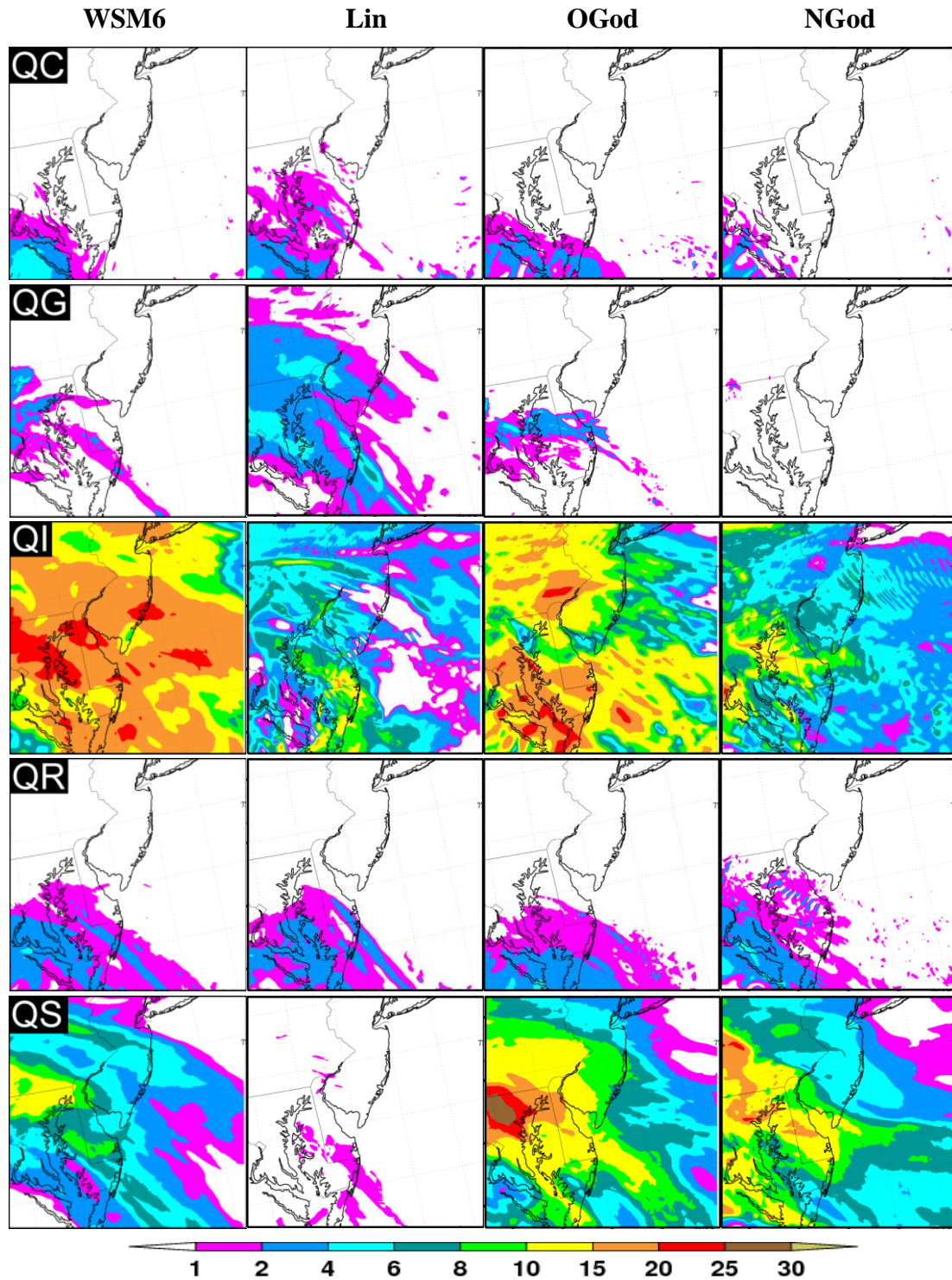


Fig. 25: Domain 4, column-integrated mixing ratios (kg m^{-2}) on 6 Feb 2010 at 06UTC. Shown abbreviations are QC = cloud water mixing ratio, QI = cloud ice mixing ratio, QR = rain mixing ratio, QG = grapple mixing ratio, and QS = snow mixing ratio. All shown values are in kg m^{-2} except QI which is scaled by 10^1 .

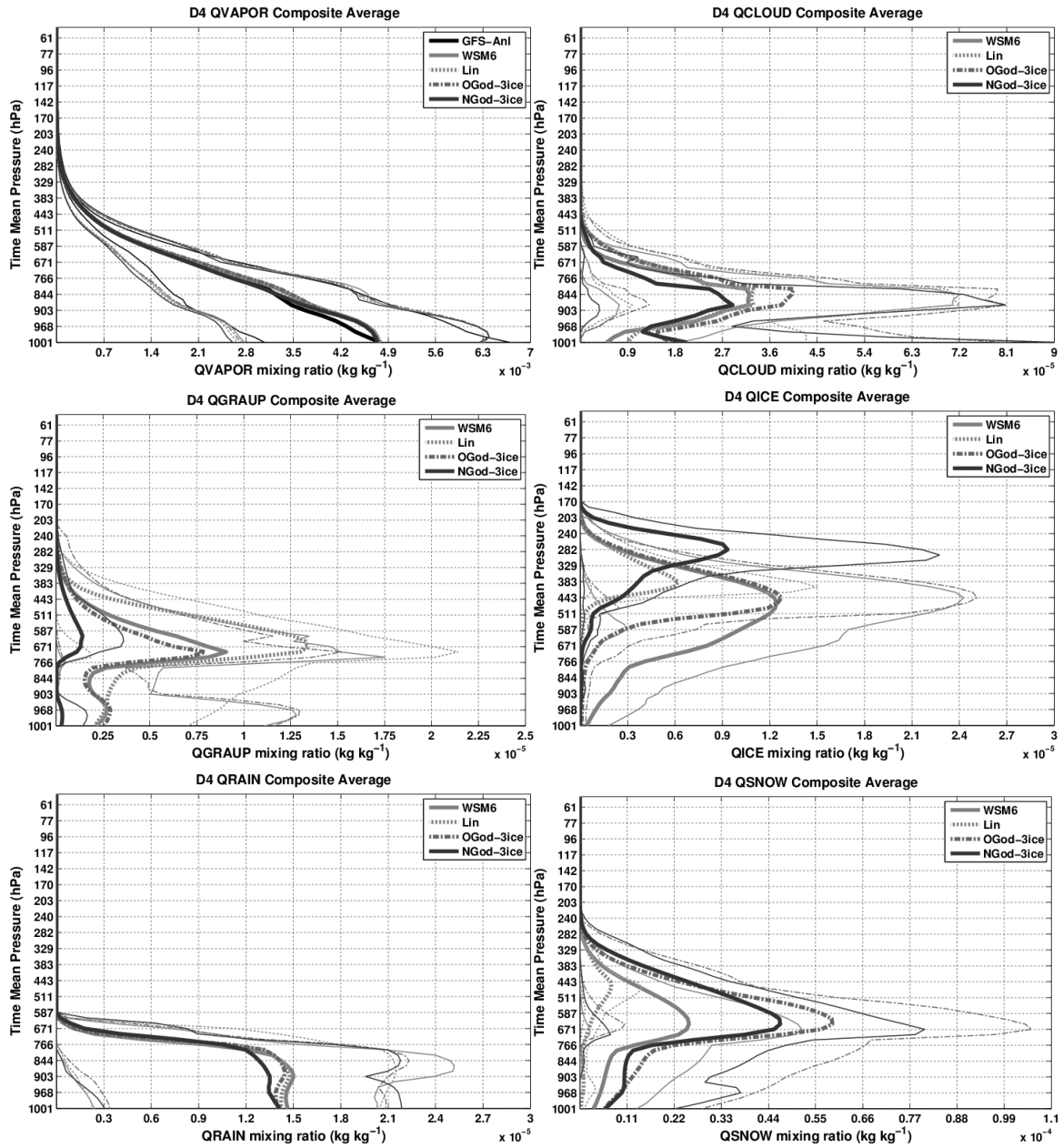


Fig. 26: Domain 4 time-averaged, case composite mixing ratios from all 6-class microphysics scheme model runs. The eight-case average mixing ratio and the cases with the highest and lowest mean mixing ratio are indicated by thick and thin lines, respectively.

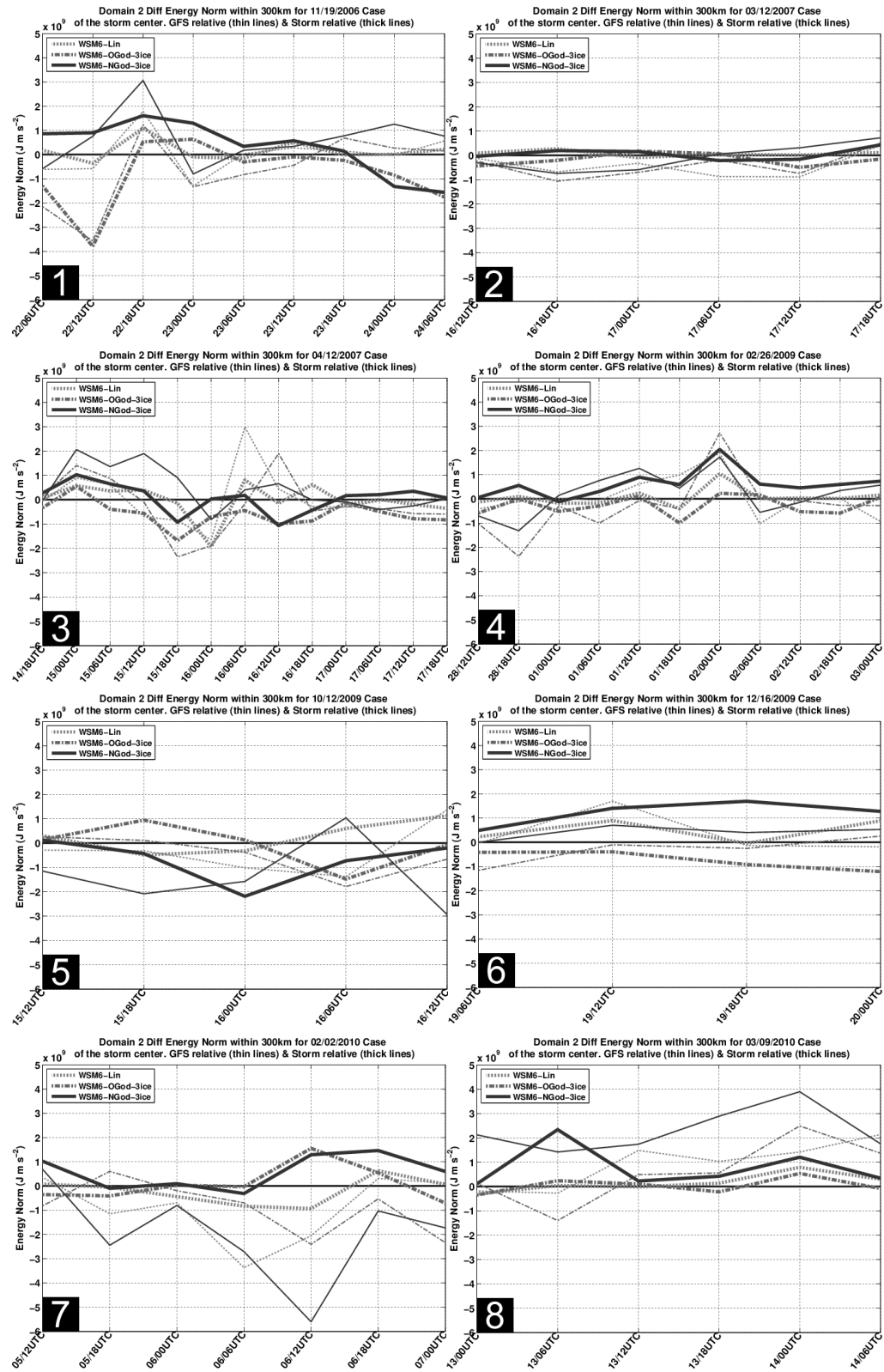


Fig. 27: Domain 2 energy norm differences for all cases within 300 km of the GMA storm center (thin lines) and each model simulated center (thick lines). Shown differences are relative to WSM6 with positive values denoting improvement.

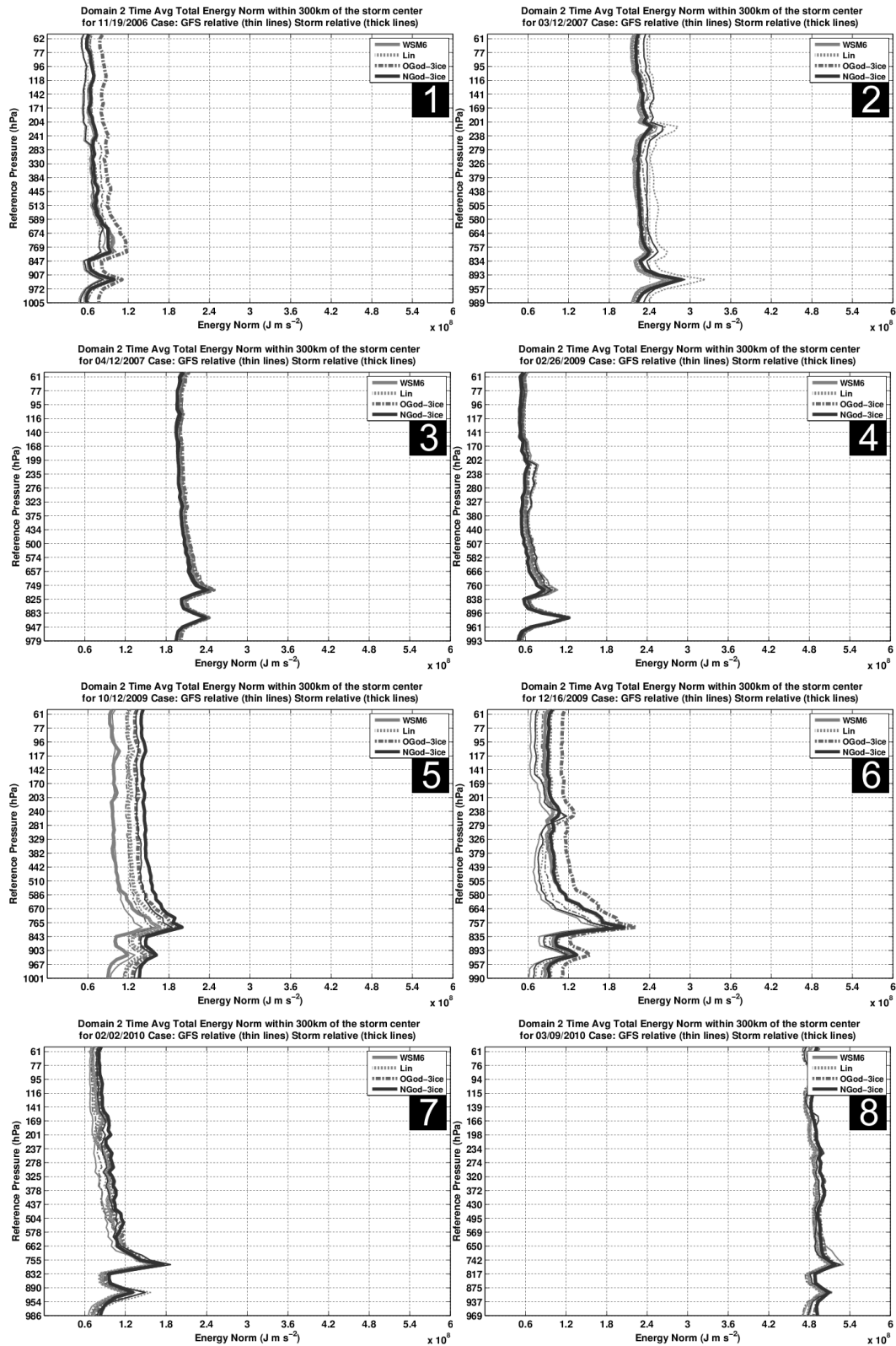


Fig. 28: 180-hr averaged domain 2 energy norm profiles for all cases within 300km of both GMA (thin lines) and each model simulated center (thick lines). The line legend for all six-class microphysics schemes is shown in the top right of each panel.

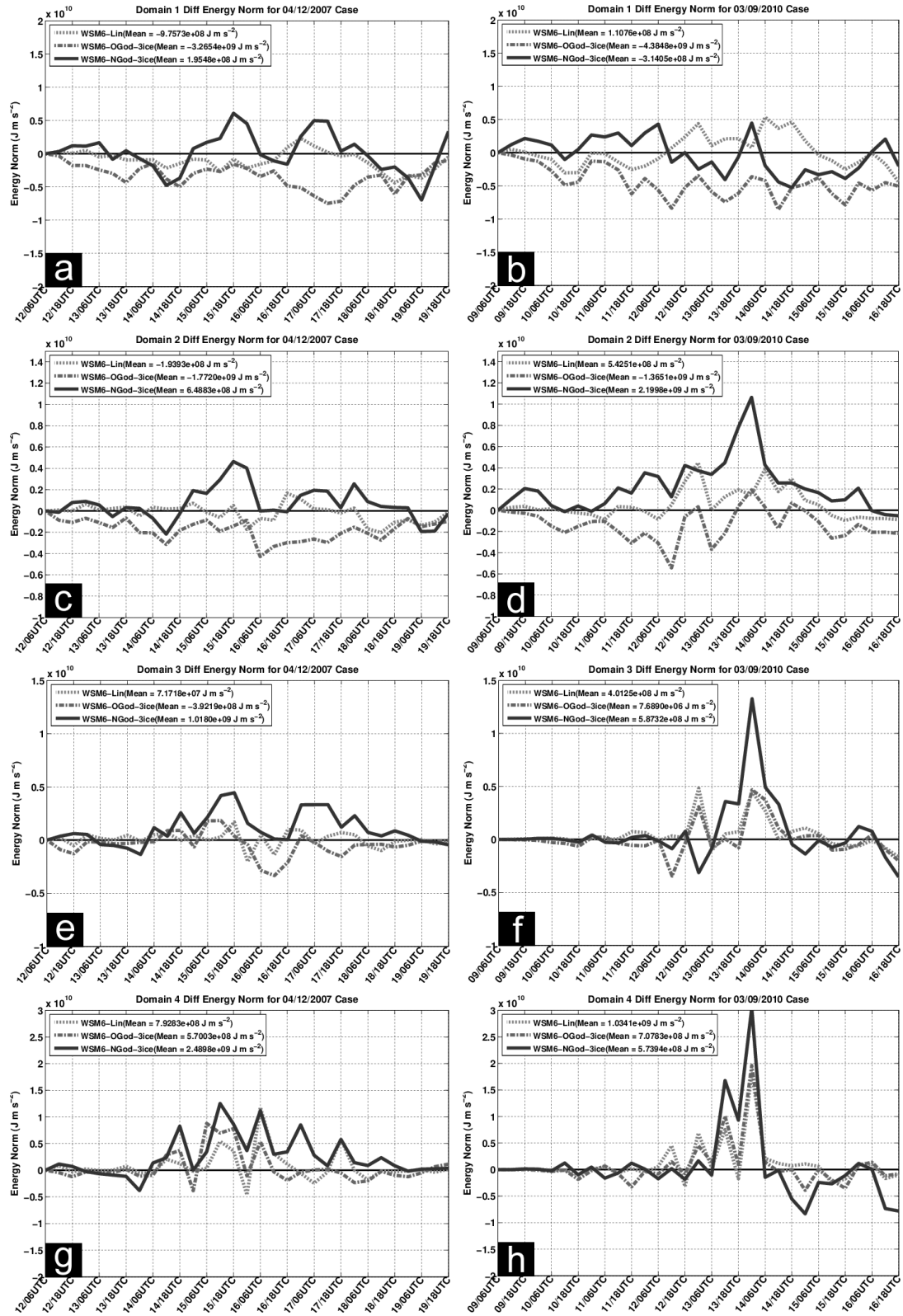


Fig. 29: Energy norm differences within the entirety of domains 1, 2, 3 and 5 for (a, c, e, g) cases 3 and (b, d, e, h) 8. Shown differences are relative to WSM6. Positive values denote improvement over WSM6.

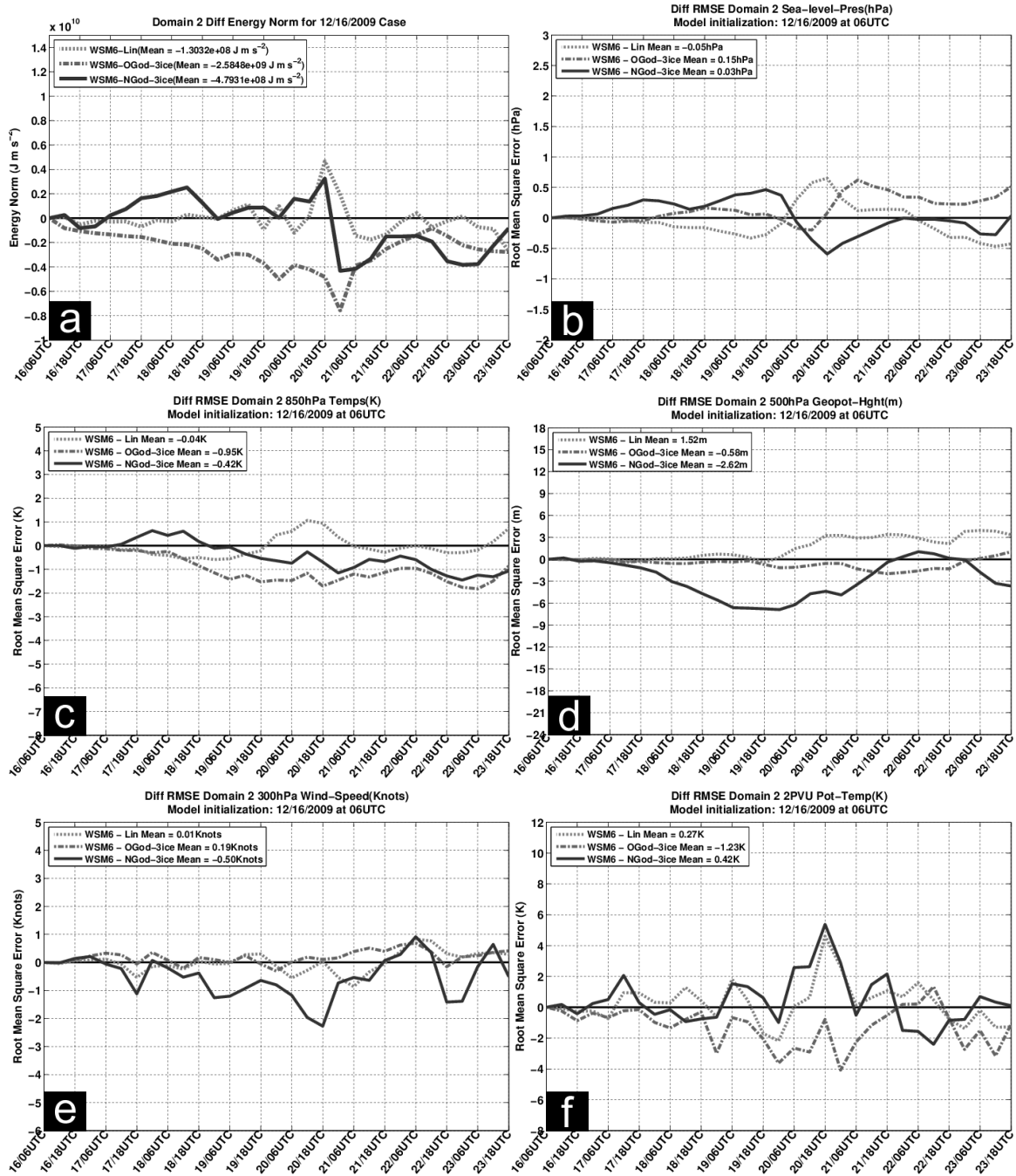


Fig. 30: Difference in domain 2 (a) energy norm, (b) sea-level pressure, (c) 850-hPa temperature, (d) 500-geopotential height, (e) 300-hPa winds, (f) 2-PVU potential temperature for case 6. Shown differences are relative to WSM6. Positive values denote improvement over WSM6.

4. Impact of cycled assimilation of radio occultation data on nor'easter simulations

4.1 Chapter abstract

The impact of Constellation Observing System for Meteorology, Ionosphere, and Climate (COSMIC)-derived radio occultation data upon Weather Research and Forecasting (WRF) Model nor'easter simulations was investigated. Simulations were conducted for 180 hours starting roughly 72 hours prior to the first precipitation impacts in the highly populated Mid-Atlantic US and associated cyclogenesis; five COSMIC periods were used (0, 24, 48, 72, and 180 hours) and a ± 1 hour observation window. For comparison, two additional model runs (± 1.5 hr COSMIC assimilation window, radiosondes only) were completed. Global Forecasting System Model Analysis (GMA) was used for verification. Many simulations (27 out of 48) exhibited a leftward bias and (30 out of 48) lagged GMA; often track error was inversely proportional to the assimilation period. Simulations with greater COSMIC assimilation exhibited an over-intensification bias due to differences in mid-tropospheric latent heating. Nor'easter-relative (within 600km of the cyclone center) and overall model simulation error decreased in upwards of 67.2% and 79.4% of six-hour periods, respectively. Model error was generally inversely proportional to assimilation period length and was sensitive to both cyclone-to-sounding distance and stratospheric data assimilation errors, yet not proportional to the total number of COSMIC observations. A majority of time periods exhibited lower model error than non-COSMIC simulation only if 48 hours of data were assimilated. Assimilation of both the wider COSMIC assimilation window and

radiosonde data led to further decreases in model error. Despite having fewer than half as many observations as radiosondes, COSMIC-related error reductions during 180 hour model simulations were still more than half that associated with radiosondes.

4.2. Methods

4.2.1. Study design

We utilized WRF Advanced Research WRF Version 3.2 (hereafter W32), which numerically solves a set of fully-compressible, non-hydrostatic, Eulerian equations in terrain-following coordinates (Skamarock et al. 2008; Wang et al. 2012). A nor'easter is a multi-scale phenomenon, which required model grids of sufficient size and resolution. The three domain grid (Fig. 31) had two-way interaction, a 45, 15, and 5 km horizontal grid-spacing, respectively, 27 vertical levels, and a 50-hPa (~20 km) top. This configuration afforded simulation of key pre-cursor synoptic and meso- α scale phenomenon (e.g., jet streaks, short- and long-wave troughs) on outer domains and smaller-scale phenomenon (e.g., orographic forcing, latent heating) on inner domains. Boundary conditions were derived from GFS forecasts ($1^\circ \times 1^\circ$ resolution).

Model simulations were conducted for 180 hours, starting roughly 72 hours prior to the first precipitation impacts in the highly populated Mid-Atlantic US and associated cyclogenesis. This time frame focused attention exclusively on cyclone initiation and its later impact in this region. A 72-hour lead time allowed simulations to spin-up, establish baroclinicity between the cooler eastern United States and warmer Gulf Stream, and simulate latent heating along the expansive (>1000 km) northern edge of the Gulf Stream. All of the above are vital for accurate nor'easter simulations (Kuo et al. 1991; Mote et al.

1997; Yao et al. 2008). Precipitation data from the New Jersey Weather and Climate Network (Robinson 2005) served as a proxy for establishing when each nor'easter first impacted the Mid-Atlantic US. Using these data, model initialization was set as 72 hours prior to the first nor'easter-related 0.5 mm (~0.02 inch) precipitation reading. A New Jersey-centric approach was chosen because of its high population density (461.6/km²), significant contribution (\$473 billion) to the US gross domestic product, and its relatively central location in the region of interest (United States Census Bureau 2012).

Model parameterizations were selected following a qualitative comparison of many tens of 12-hour, single-domain WRF simulations during a November 2006 nor'easter to GFS model analysis (GMA). The selected parameterizations (see below) were from the model run providing both the best comparison to GMA and the ability to complete a three-domain (see Fig. 31), 180-hour model simulation in less than 12 hours of computational time.

- Longwave radiation: RRTM (Mlawer et al 1997)
- Shortwave radiation: Dudhia (Dudhia 1989)
- Microphysics: Goddard 3-ice, graupel (Lang et al. 2007)
- Surface layer: MM5 similarity (Zhang and Anthes 1982)
- Land surface: NOAH (Chen and Dudhia 2001)
- Boundary layer: BouLac (Bougeault and Lacarrère, 1989)
- Cumulus parameterization: Grell-Devenyi ensemble scheme (Grell and Devenyi 2002)

This study focused on eight nor'easter cases (Table 10) selected based upon their timing, intensity, and track. All cases occurred after the April 2006 launch and

deployment of COSMIC (Anthes et al. 2008) and during the “nor’easter season” (October to April) as defined by Jacobs et al. (2005). To make this study more generalized, we did not select only the most severe nor’easters, but instead aimed for a relatively small but diverse sample. Case severity was measured using the Northeast Snowfall Impact Scale (NESIS; Kocin and Uccellini 2004). Under NESIS, nor’easters were classified on a scale of 1 (notable) to 5 (extreme) based upon the population impacted, area affected, and event severity. Finally, all cases were required to have precipitated in New Jersey.

To understand how COSMIC impacted nor’easter simulations, one non-assimilation run and five COSMIC assimilation runs (varying assimilation duration) were completed for each event. The five COSMIC runs focused on the pre-coastal transition phase and varied only by their assimilation period length: 0 (only at initialization), 24, 48, 72, and 180 (full assimilation) hours. Despite its simplicity, we used cycled, three-dimensional variational data assimilation to assimilate COSMIC into WRF (Barker et al. 2004; Skamarock et al. 2008; Huang et al. 2009) rather than ensemble Kalman filter or four-dimensional variational data assimilation for consistency with parallel research work. To match both radiosonde data and GFS boundary conditions, COSMIC data were assimilated every 3 hours utilizing a ± 1 hour data window. All COSMIC observations were cut-off at 20 km to minimize stratospheric data assimilation errors (Cucurull et al. 2008). The assimilation background error covariance matrix was derived from a series of 24-hour, single-domain (domain 1) WRF runs initialized every 12 hours during January 2010. Additional model runs assimilating all COSMIC data (i.e., ± 1.5 hour window) and radiosonde data for all 180 hours were completed for comparison purposes.

4.2.2. Verification and analysis techniques

Simulation validation data was derived from the GFS model analysis (hereafter, GMA). We favored these data because all three WRF domains included data sparse regions where in-situ observations were often not available, and GMA was easily interpolated to the WRF model grid. Furthermore, the lateral boundary conditions were GFS-based to prevent additional data from being assimilated indirectly via GFS model analysis at post-WRF initialization.

Model run analysis was comprised of several parts. Analysis increments evaluated how COSMIC assimilation perturbed dynamical fields. Storm tracks were determined using local-minima in sea-level pressure (SLP) via an objective, self-developed algorithm similar to that used at the Climate Prediction Center (Serreze 1995; Serreze et al. 1997). Quantifying storm track simulation differences included finding the minimum storm SLP and GMA-relative track bias. Overall accuracy of the local-storm environment (i.e., within a 600 km wide, WRF-centered box) and large-scale environment (i.e., entire model domain) was evaluated using the dry energy norm (Rabier et al. 1996):

$$\langle \mathbf{X}, \mathbf{Y} \rangle = -\frac{1}{2} \int_{p_{sfc}}^{p_{top}} \iint_A \left(\Delta u^2 + \Delta v^2 + R_d T \Delta \ln(p_{sfc})^2 + \frac{c_p}{T_v} \Delta T^2 \right) dx dy dp \quad (2)$$

In (2), \mathbf{X} is the WRF model state vector, \mathbf{Y} is the GMA state vector, u is the zonal wind (m s^{-1}), v is the meridional wind (m s^{-1}), R_d is the dry air gas constant ($287 \text{ J kg}^{-1} \text{ K}^{-1}$), T_r is the mean surface temperature (K), P_{sfc} is the surface pressure (Pa), c_p is heat capacity at constant pressure ($1004 \text{ J kg}^{-1} \text{ K}^{-1}$), and T is air temperature (K). A 600-km wide box was used to evaluate the local-storm environment because it captured the storm, yet minimized background environment contamination. Point-to-point root mean square

error (RMSE) calculations were performed for sea-level pressure (SLP), 850-hPa temperature, 500-hPa geopotential height, 300-hPa winds, and 2 potential vorticity unit (2 PVU; $1 \text{ PVU} = 10^{-6} \text{ K m}^2 \text{ kg}^{-1} \text{ s}^{-1}$) potential temperature. The first four variables were frequently referenced in Kocin and Uccellini (2004) for nor'easter analysis, and 2-PVU potential temperature was used to investigate changes to the dynamic tropopause. For both the energy norm and RMSE, smaller values denote less error.

The energy norm and RMSE metrics served complimentary purposes. First note that the energy norm is a volume integration, whereas RMSE is a layer integration. Thus the former better represents the entire model simulation and was less sensitive to large errors in single layers. Second, the energy norm involves four variables (surface pressure, temperature, zonal wind, and meridional wind) and not just one, making its results more robust. Buizza et al. (2005) provided a compelling argument for the use of the energy norm as the primary validation metric and its current usage at ECMWF for model validation. Given all the above, the energy norm was the primary vehicle to evaluate model version simulation quality, and RMSE helped to identify error sources at levels throughout the troposphere.

4.3. Results

4.3.1. COSMIC observations and analysis increments

For each assimilation time and case, the number of assimilated observations and analysis increment magnitude varied notably. Figure 32 shows the number of COSMIC observations assimilated in domain 1 as a function of time and their mean. As seen in Fig. 32, assimilation-to-assimilation variability ranged from 0 (case 1) to 33 (case 3)

observations per three hours. Unlike later cases, the first three cases exhibit higher observation variability and a roughly 24-hour periodicity which is attributed to the deployment of COSMIC itself. As noted in Anthes et al. (2008), the six micro-satellites were launched as one unit and were then slowly adjusted to their eventual 30° longitude separation over a period of 18 months. Therefore later cases (after 2008) displayed lower observation variability, lacked any 24-hour periodicity, yet the mean number of observations per cycle was roughly conserved.

Figure 33 depicts 500-hPa geopotential height, SLP and 850-hPa temperature and water vapor analysis increments for 12 UTC 24 Nov. 2006 and 18 UTC 2 Mar. 2009. These two times were selected due to their large analysis increments. As seen in Fig. 33, analysis increments are tightly linked to observation location, but their severity varies from point to point. As an example, Fig. 33b shows an observation in the top right corner which produced an over 30 m rise in 500-hPa geopotential height, yet another location 500 km northeast of Puerto Rico in Fig. 33a resulted in a near 0 m change. Similar findings were noted when COSMIC observations were more clustered (e.g., Fig. 33a in Kansas compared with Fig. 33c in Northern Mexico). Therefore, analysis increments are not proportional to the number of observations, but instead to the degree of deviation between the COSMIC sounding and the WRF background environment.

4.3.2. WRF-COSMIC run analysis

Figure 34 displays SLP-based storm tracks for all simulations. As seen in Fig. 34, storm track variability ranged from relatively minor (case 5) to extreme (case 4). To quantify this variability and discover potential biases, Fig. 35 displays GMA-relative

track errors every six hours (small symbols) and their mean (larger symbols) for each WRF run. Shown track errors are not latitude/longitude based, but are instead aligned to GMA cyclone propagation direction every six hours which defines the positive y-direction. In this GMA-relative framework, WRF track error ranged from around 50 km (case 3) to upwards of 2,100 km (case 4), and was typically within 150-230 km. Most WRF model runs (5 out of 8 cases) exhibited a leftward track bias (100 - 200 km) or lagged (50-200 km) GMA. To further quantify these biases, Table 11 indicates the total number of six-hour periods where each individual model simulation exhibited a particular track bias relative to GMA. With the exception of WC-72hr run (72-hour assimilation period), Table 11 indicates all simulations favored a leftward track bias or lagged GMA at least 51.2% (WC-48hr) and as much as 60.3% (W32) of the time. As seen in Table 11, simulations assimilating COSMIC data for longer periods, became increasingly likely to either exhibit a rightward track bias that lagged GMA or a leftward track bias that led GMA, but were decreasingly likely to both lag GMA and exhibit a leftward track bias. This change in track error tendency is attributed to assimilated COSMIC observations typically bringing simulations closer to GMA (up to 500 km; case 4). Assimilating these data however was insufficient to overcome the inherent biases of W32 for each case. For all cases, the W32 run and WC-0hr run (1 assimilation cycle) storm track and its error are nearly indistinguishable. Generally mean track error was found to decrease with increased COSMIC assimilation, but there were exceptions. These exceptions were likely caused by compensating track errors (Fig. 35, case 8) or simulated field errors (Fig. 35, case 2). For the former case, W32 has the same degree of average track error as WC-180hr, however closer inspection reveals W32 exhibited a higher degree of track error

than WC-180hr. For the latter case, Fig. 35 shows WC-48hr run (48 hour assimilation) track errors averaged over 600 km, much higher than all other runs. As seen in Figs. 36a and 36b, WC-48hr-based 500-hPa geopotential height (contours) and relative vorticity (positive only, shaded) were both displaced westward relative to GMA. This westward displacement likely caused the simulated surface low to track over western New York rather than off the East Coast.

To demonstrate how cycled data assimilation impacted nor'easter simulations, Fig. 36 displays various dynamical fields at key times for case 2 (Figs. 36a-c) and case 4 (Figs. 36d-i). These particular times were selected because all model runs had generated a surface low and simulation differences were readily noticeable. For case 2, all model runs generated the initial surface low too far north (Fig. 34) resulting in an overall leftward model track bias (Fig. 35). Figure 36a-c provides an explanation in the form of an over-sharpened geopotential height trough near Lake Michigan and the severe underestimation ($>10^{-4} \text{ s}^{-1}$) of the vorticity maximum situated over Southern Georgia. These errors enhanced positive vorticity advection over Lake Michigan and decreased it over Georgia which led to enhanced and weakened ascent, respectively. Another possible explanation was the dearth of COSMIC observations within 300 km of either feature within 48 hours prior to this time (not shown).

Figs. 36d-f and Figs. 36g-i display 500-hPa geopotential height and SLP for GMA, W32, and WC-180hr on 18 UTC 28 Feb. 2009 and 18 UTC 2 March 2009, respectively. These times were selected to emphasize storm initialization and its development 48 hours later. On 28 Feb. all shown model data is nearly identical, yet by 2 March, they differed radically as was shown in Fig. 36. Eventually, these differences

resulted in a 72-hour time lag with an 1,800 km (Fig. 35) track error between W32 and GMA, which was attributed to errors with the 500-hPa long-wave trough situated over the eastern US. Figures 36g-i demonstrate that both W32 and WC-180hr simulate a prolonged cut-off geopotential height minimum near Florida, yet the cut-off existed less than one day in GMA. Due to this cut-off, propagation speed sharply fell under 5 m s^{-1} , whereas GMA had the system moving rapidly northeastward, resulting in the extremely large track errors seen in Fig. 35d. Due to COSMIC assimilation, the cut-off height minimum remerged with the poleward 500-hPa long-wave trough roughly 24 hours sooner than W32.

To assess how COSMIC influenced simulated cyclone intensity, minimum SLP was compared between each simulation and GMA. Simulated SLP was deemed “accurate” if it was within 5 hPa of GMA. Overall, WRF model simulations over-intensified a nor’easter upwards of 14.0 hPa (case 4; WC-180hr) and under-intensified it by as much as 18.3 hPa (case 6; WC-180hr), which equate to 1.93 and -1.47 times the standard deviation of 9.50 hPa, respectively. For COSMIC runs, simulated minimum SLP, except the 72 hour assimilation run (WC-72hr), were accurate in at least 4 of 8 cases. Over-intensification was slightly favored for both WC-72hr and WC-180hr runs, where 3 and 4 out of 8 cases were too intense, respectively. The other COSMIC runs did not exhibit any definitive intensification bias. Investigations into these intensification tendencies included analysis of various meteorological fields (i.e., temperature, relative vorticity, latent heating, etc.) which attributed storm intensity variations to differences in mid-tropospheric latent heat release and consequently upward vertical motion. As an example, Fig. 37 displays 500-hPa latent heating and 300-hPa winds with SLP overlaid

on 15 Oct. 2009 at 12 UTC and 18 UTC. These times were selected because the W32 and WC-180hr simulations formed the nor'easter in close proximity ($<50\text{km}$ separation) and had nearly identical meteorological fields at 12 UTC. As seen in Fig. 37, both simulations display strong latent heat release at 500hPa and position the cyclone in the right-exit region of a 300-hPa jet. Six hours later, minimum SLP fell 5 hPa and 4 hPa in the WC-180hr and W32 runs, respectively. At 18 UTC local changes in geopotential height, temperature, and relative vorticity (not shown) were comparatively small, the cyclone remained in the right-exit region of the 300-hPa jet, and latent heating remained strong along the cyclone track. Despite such similarities, simulated latent heat release was stronger for the WC-180hr run and resulted in greater SLP falls in WC-180hr than in W32. For all other cases and times, latent heat release often acted in concert with other factors (e.g., positive vorticity advection) to cause simulated SLP changes, but SLP falls were nearly always co-located with the region of strongest latent heating.

Local-storm environment accuracy was evaluated via energy norm differences (W32 - model) for both the GMA- and WRF-centered frameworks in Fig. 38. Both frameworks include the same 600-km-wide box except one is GMA-centered and the other is centered upon where each simulation positioned each nor'easter, respectively. In Fig. 38, thin lines represent GMA-centered and thick lines represent WRF-centered energy norm differences. Because the W32 energy norm is subtracted, positive values denote improvement versus W32. As a companion to Fig. 38, Table 12 indicates the number of six-hour periods where COSMIC runs had a lower energy norm than W32 in each framework. Figure 38 and Table 12 indicate COSMIC assimilation improved upwards of 41 out of 61 (67.2%) WRF-centered periods, yet only 29 out of 61 (47.5%)

GMA-centered periods exhibited improvement. In order of increasing assimilation, WRF-centered simulations were improved more than 50% of the time in 3, 3, 6, 4, and 6 cases. This trend is likely non-linear due to both sounding location and small sample size. For instance, the WC-72hr run improved exactly 50% of the six-hour periods in cases 4 and 5. For the former case, no soundings were present within 800 km of the cyclone center during the last 18 hours of the data assimilation period, potentially leading to simulation errors. For case 5, only 50% of periods were improved because there were only four valid time steps and not because of substantial model error. In contrast, COSMIC runs reduced GMA-centered energy norms in greater than 50% of the periods in 3 out of 8 cases. This result is largely explained by storm-position error as illustrated best by Fig. 36 which displays large simulated field differences between WRF runs and GMA. A common result to both frameworks was the poor showing of the WC-24hr run, which was attributed to both the WC-0hr and W32 runs being nearly identical and possible data assimilation errors related to model spin-up.

To identify potential energy norm sources, WRF- and GMA-centered energy norms were time-averaged and horizontally-integrated at each vertical level (Fig. 39). While Fig. 39 shows energy norm contributions from all levels, the most distinctive error sources originated at four levels: ~850 hPa, ~600 hPa, ~300 hPa, and ~180 hPa. Further analysis (not shown) attributed these energy norm spikes to errors in the planetary boundary layer height and low-level jet, the intensity and positioning of a mid-tropospheric jet, jet stream intensity and positioning, and tropopause height, respectively. Increased error near the model top was due data assimilation errors propagating from the stratosphere (Cucurull et al. 2008). As seen in Fig. 39, the largest energy norm

contributor in both GMA- and storm-relative environments exists at 650 hPa. The 650-hPa jet is driven by height and in turn temperature gradients and therefore a possible link may exist to differences in latent heat release similar to that seen for SLP in Fig. 37.

Energy norm differences for the entirety of domains 1, 2 and 3 were used to assess the broader implications of COSMIC assimilation. Figure 40 displays these differences for cases 1 and 4. For both cases the WC-180hr run exhibited the least error, yet only in case 1 was the energy norm seemingly inversely proportional to assimilation period length. This result was attributed the notable difference in sounding-to-cyclone distance for these two cases. Specifically, all case 1 runs had at least one observation within 200 km of the cyclone center at the end of each assimilation period, whereas in case 4 this separation was well over 800 km until 72 hours into the simulation. At 72 hours a COSMIC observation was taken over Northwest Georgia within 100 km of the cyclone center which alone reduced SLP 0.5 hPa and more importantly lowered 500-hPa geopotential height 30 m (not shown). After this observation, energy norm differences for domains 2 and 3 (Figs. 40d and 40f) remained positive for both the WC-72hr and WC-180hr runs. With the exception of the domain 1, principal energy norm sources on the entire domain (not shown) were commensurate with those from the local-storm environment analysis. Due to the larger spatial extent of domain 1 as compared to the local-storm analysis coverage, the energy norm contribution at 300 hPa was greatly enhanced due to the upper-level jet covering a large fraction of domain 1. Due to its large spatial coverage (approximately 1,000-3,000 km long and 500-1000 km wide), intensity ($\sim 75 \text{ m s}^{-1}$) even small positional errors associated with the upper-level jet produced notable error. In contrast, because the 600-hPa jet is both smaller (500-1000 km long and

300-500 km wide) and weaker ($< 40 \text{ m s}^{-1}$) its contribution to the domain 1 energy norm is greatly reduced for domain 1. Energy norm differences for all of domains 2 and 3 are entirely consistent with the local-storm analysis largely due to their smaller size and more southerly positioning. As a result, the upper-level jet was either not present or covered a small fraction of either domain, whereas the 600-hPa jet associated with the nor'easter tended to cover a larger fraction of these domains.

Nor'easter simulations benefited from COSMIC assimilation in at most 93 (37.5%), 207 (83.5%), and 197 (79.4%) out of 248 six-hour periods in domains 1, 2, and 3, respectively (Table 12). Similar to the local-storm environment analysis, WC-24hr simulations commonly exhibited greater error. Excluding WC-24hr runs, the number of improved periods was largely proportional to the length of COSMIC assimilation unlike the local-storm environment analysis. In order of increasing assimilation duration, greater than 50% of six-hour periods domain 1 had 0, 0, 0, 2, and 3 cases improved, domain 2 had 3, 3, 4, 7, and 8 cases improved, and domain 3 had 3, 3, 4, 7, and 8 cases improved. Of all eight cases, case 2 exhibited the least improvement due to large sounding-to-cyclone distances ($> 800 \text{ km}$) and errors in 500-hPa geopotential height fields over the northeastern United States (Fig. 36).

For comparison to the energy norm results and to determine which meteorological fields were most sensitive to COSMIC assimilation, RMSE calculations were completed and summarized in Table 12. The five RMSE error variables shown in Table 12 were selected either due to their strong emphasis in Kocin and Uccellini (2004) or as an indicator of the dynamic tropopause accuracy. Similar to the energy norm, values shown in Table 12 reflect the number of six-hour periods where a particular model has a lower

RMSE value than the non-assimilation WRF run. In comparison to the energy norm, domain 1 RMSEs indicated at least twice as many periods and RMSEs for domain 2 were consistent with the energy norm. The one exception for all domains was 2-PVU potential temperature. Increased error near the model top (Fig. 39) suggests this exception was caused by compounding errors originating from upper troposphere temperature and wind errors or the high-degree of fine-scale variability associated with this field, or propagating stratosphere-based data assimilation errors as suggested by Curcurull et al. (2008) or a combination of the above. For the other four RMSEs, data assimilation periods of 72 hours or greater were beneficial at least 50% of the time for all domains, and periods longer than 48 hours for domains 2 and 3. To illustrate how RMSEs compared to the energy norm, Fig. 41 shows all of these error metrics for case 3, domain 2. Figure 41 infers that although each metric varied with respect to W32 (i.e., the zero line on each plot), the relationship between each of the shown COSMIC runs was approximately conserved.

Given the positive RMSE results for WC-180hr, several speculations were made about the potential energy norm sources in domain 1. First, from Table 12, SLP and 2-PVU potential temperature RMSEs indicated less improvement as compared to the other three RMSE metrics. As suggested earlier, there is a strong association between latent heat release to cyclogenesis, convection, and upward vertical motion in the troposphere and by association SLP tendency as illustrated by Fig. 37 and suggested by Kocin and Uccellini (2004). Increased SLP errors would correspondingly increase the surface pressure energy norm component. Second, model error was higher near the stratosphere and was likely related to the greater northern extent (~3,000-4,000 km) of domain 1. In

domain 1, a greater proportion of the model volume existed in the stratosphere and therefore possibly leading to increased error. Finally, only domain 1 directly interacted with the GFS model boundary conditions, which itself also had errors and contributed to the energy norm.

4.3.3. Comparison to larger COSMIC assimilation window and radiosondes

To put the above results in context, model runs assimilating both COSMIC data with a ± 1.5 hour window (hereafter WCL-180hr) and radiosondes (hereafter WR-180hr) were compared to WC-180hr and W32. A wider assimilation window permitted comparison between the positive benefits gained from additional observations and the consequences of increased temporal error. Radiosonde assimilation afforded comparison between COSMIC and the well-established and tested radiosonde network. For both new datasets, the same 3-hour assimilation cycle for WC-180hr simulations was used.

Differences in window size and usage of radiosondes led to considerable variation in the quantity of data assimilated for a given time. Overall, WC-180hr averaged 12.74 soundings per three hours and accumulated anywhere from 679 (case 5) to 847 (case 7) observations throughout the entire run. Adding 30 minutes to the assimilation window size raised the COSMIC data assimilation average to 18.65 soundings per three hours and total assimilated observations ranged from 989 (case 5) to 1,265 (case 7) observations for the entire run. Finally, the WR-180hr run averaged 34.5 observations per three hours and 2,070 observations total. Because radiosondes are launched twice daily, with the exception of special launches, some assimilation times had no corresponding radiosonde

data. Regardless, WR-180hr still assimilated nearly twice as many soundings overall, owing to its numerous fixed locations.

Figure 42 displays GMA-relative track errors every six hours (small symbols) and their mean (larger symbols) for each WRF run. Shown track errors range from around 50 km (case 5) to 2000 km (case 4) and averaged less than 300 km. Storm tracks favored a leftward bias in 19 out of 32 (59.3%) simulations averaging between 50-150 km and four cases overall. Notably, the WR-180hr run exhibited a leftward track bias in two additional cases (cases 6 and 8). Further analysis showed that stronger mid-tropospheric latent heat release was the cause (not shown). To further quantify these biases, Table 13 displays the total number of six-hour periods where each of the 180-hour model runs and W32 exhibited a particular track bias relative to GMA. With the exception of WCL-180hr run, all simulations favored a leftward track bias or lagged GMA in at least 51.7% (WC-180hr) and as much as 67.8% (WR-180hr) time periods. Of all simulations, WR-180hr runs exhibited notably stronger leftward track bias (10 more time periods) than any other run which was attributed to its observations correcting meteorological fields primarily over land. As a result, WR-180hr nor'easter simulations tended to track closer to land, and thus typically leftward of GMA, W32 and both COSMIC runs (especially cases 2 and 6). For WCL-180hr runs 58.1% of time periods exhibited a rightward track bias relative to GMA, but similar to WC-180hr simulations, a majority of those time periods (55.6%) lagged GMA. Despite its rightward track bias, WCL-180hr simulations did not vary notably from WC-180hr simulations. Instead the rightward tendency of WCL-180hr originates from its additional three time periods in case 6 and the weak synoptic pattern in case 8 which together account for 6 of 9 rightward time steps not exhibited by WC-

180hr run. With the exception of case 8, WCL-180hr storm tracks were more accurate and precise than WC-180hr simulations. For this case, a 500-hPa geopotential height trough in WCL-180hr was approximately 50 km further west, which resulted in the surface low tracking more in land and further from GMA. For other cases, WCL-180hr simulations had anywhere from 0 km (case 6) to 1,400 km (case 4) less error than WC-180hr at a given time. The notable improvement in case 4 was attributed to the additional COSMIC soundings over the northeastern Gulf of Mexico which helped WRF to remerge the 500-hPa geopotential height cut-off (Fig. 36) into the poleward longwave trough 24 hours sooner than in WC-180hr simulations. Impressively, the WR-180hr run decreased the track error an additional 300 km as compared to WCL-180hr owing to numerous radiosonde stations in the southeastern US.

Cyclone intensity accuracy relative to WC-180hr was then evaluated. From the previous section, WC-180hr cyclone intensity was accurate in 4 out of 8 cases and was over-intensified in 3 out of 8 cases. As compared to WC-180hr simulations, WCL-180hr and WR-180hr altered storm intensity up to 7.4 hPa and 22.7 hPa, respectively. Case-wise, WCL-180hr was accurate in 5 out of 8 cases, over-intensified in 2 out of 8 cases, and better simulated intensity in 6 out of 8 cases versus WC-180hr simulations. Impressively, WR-180hr cyclone intensity was accurate in 7 out of 8 cases and unlike either COSMIC runs, no over-intensification occurred. As noted in the previous section, cyclone intensity accuracy is linked to mid-tropospheric latent heat release and the quantity of observations near the cyclone. These two factors improved the WR-180hr simulation of cyclone intensity by 22.7 hPa (case 6) as compared to WC-180hr. For this case, numerous radiosonde stations along the immediate US east coast helped the WR-

180hr run generate a strong region of mid-tropospheric latent heating along a frontal zone, which in turn lowered SLP. In contrast, COSMIC soundings during cyclogenesis for this case were over the Pacific Ocean and western US which led to weaker 500-hPa latent heat release for both COSMIC simulations. As compared to WC-180hr simulations, cyclone intensity was more accurately simulated for 7 out of 8 and 4 out of 8 cases in WCL-180hr and WR-180hr simulations, respectively. Therefore, although radiosonde assimilation improves cyclone intensity in these simulations more so than COMSIC, the ability of COSMIC to gather data from remote regions is still quite beneficial.

Figure 43 displays simulation energy norm differences ($W32 - \text{model}$) in both GMA (thin lines) and WRF-centered (thick lines) frameworks. As in Fig. 38, the energy norm was integrated within a 600 km wide volume only when all simulations had the nor'easter. Comparing energy norm differences reveals WRF-centered energy norms were lower than their GFS-centered counterparts in 46 out of 60 (76.6%) periods. Thus, full data assimilation runs had less error simulating the local-storm environment than that related to track and position error. For quantitative comparison, Table 14 indicates the total number of six-hour periods where the energy norm for each run was less than W32. As inferred from Table 14, WCL-180hr and WR-180hr WRF-centered simulations were improved over WC-180hr for an additional 3 and 13 periods, respectively. Gains were more impressive in GMA-centered environment simulations. In particular, the energy norm was lower in WCL-180hr and WR-180hr for an additional 8 and 13 time periods than WC-180hr. Direct comparison to WC-180hr, revealed WCL-180hr and WR-180hr bested WC-180hr in at least 41 out of 60 (68.3%) and 48 out of 60 (80%) periods,

respectively. These results suggest that storm- and GMA-centered simulations benefitted from a wider assimilation window associated from WCL-180hr simulations increased temporal error from assimilated COSMIC observations. Additionally, despite nor'easters typically developing offshore, simulations of their local environment benefitted more from radiosondes than COSMIC assimilation.

Next, to evaluate how these assimilated data influenced WRF, energy norm integration was expanded to include all times and domains. To demonstrate how energy norm differences varied with model domain, Fig. 44 displays these differences from cases 1 and 4. As seen in Fig. 44, WR-180hr energy norm differences are typically more positive than either COSMIC run, and WCL-180hr is consistently better than WC-180hr only for domains 2 and 3. For quantification purposes, Table 14 indicates for how many periods the energy norm is lower than W32 for each assimilation run. Results from domains 2 and 3 indicate model assimilation in general benefitted WRF simulations during at least 197 of 248 (79.4%) periods. For these two domains, WCL-180hr exhibited little difference (3 period variance), whereas WR-180hr was notably better (24 additional periods [$\sim 9.6\%$ improvement]) in comparison to WC-180hr. Directly comparing the WR-180hr and WCL-180hr energy norms to WC-180hr revealed larger variability amongst assimilation runs. Specifically, 164 (66.3%) and 227 (91.5%) of all six-hour periods have lower energy norms in WCL-180hr and WR-180hr, respectively. Domain 1 energy norm results varied radically from domains 2 and 3. As seen in Table 14, WC-180hr and WCL-180hr energy norms were lower than W32 in only 93 (37.5%) and 82 (33.1%) six-hour periods, respectively. Despite similar comparison to W32, WCL-180hr had a lower energy norm than WC-180hr in only 66 (26.6%) periods. Analysis of the energy norm at

all vertical levels (not shown) revealed WCL-180hr simulations experienced greater error toward the model top due to it assimilating more observations from poleward regions where stratospheric data assimilation errors were more likely. In stark contrast to the COSMIC runs, Table 14 shows radiosonde assimilation to be largely beneficial. Specifically, WR-180hr energy norms were lower than W32 in 184 (74.2%) and WC-180hr in 223 (89.9%) periods. In comparison to domains 2 and 3, energy norm results for domain 1 improved 43 (17.3%) fewer periods. Similar to COSMIC, WR-180hr simulations also exhibited increased error near the model top; however, fewer radiosonde stations exist at more poleward latitudes which limited error growth. Two-tailed t-tests revealed energy norms from each set of full assimilation runs were significantly different (p-value ~ 0) than W32, but only WR-180hr was significantly different (maximum p-value = 0.0023) than WC-180hr.

Figure 45 displays domain 2 energy norm and 5 RMSEs differences from case 3 and Table 14 quantifies simulation improvement as compared to W32 for all cases and domains. As seen in Fig. 45, RMSEs from WR-180hr are generally lower than both COSMIC runs, but there are exceptions (e.g., Fig. 45b). Table 14 shows that with the exception of SLP RMSE, WC-180hr improved more periods relative to W32 than WCL-180hr. This improvement, however, was relatively modest (not more than 11 [4.4%] periods) and the differences qualitatively were small (Fig. 45). In contrast, WR-180hr simulations trumped both COSMIC runs in all RMSE metrics except domain 3 2-PVU potential temperature by no fewer than 35 (14.1%) additional periods. Once again, 2-PVU potential temperature RMSE results varied substantially (>139 [52.0%] period difference) from other RMSE results in all runs. The consistently poor 2-PVU potential

temperature RMSE results underscore the likelihood that all assimilated sounding data are subject to error propagation from the stratosphere.

4.4. Conclusions

An investigation into the impact of COSMIC assimilation upon eight nor'easter simulations was conducted. As part of this investigation, both a data denial experiment and comparison of full-length (all 180 hours) data assimilation runs were completed. Data denial experiments involved assimilating COSMIC through five assimilation periods (0, 24, 48, 72, and 180 hours) and focused particularly on the critical incipient stages of nor'easters. For the full-length data assimilation comparison WC-180hr was compared to both WCL-180hr which had a larger assimilation window and WR-180hr which assimilated radiosondes. All simulations ran for 180 model hours, were initialized 72 hours prior to the first precipitation impacts in the highly populated Mid-Atlantic US, assimilated observations via cycled, three dimensional data assimilation, and only differed according to their assimilation period length or window size, or data source. Model accuracy was assessed by comparing each WRF run to GMA.

Analysis of SLP-based storm track demonstrated notable variability amongst the model runs and between cases. Track errors were not always proportional to assimilation period length, especially when soundings were far removed (>300 km) from the cyclone which particularly impacted cases 2 and 4. In these two cases, WRF simulated nor'easters were displaced more than 800 km relative to GMA due to inaccurate simulations of 500-hPa relative vorticity advection and latent heating and 500-hPa geopotential heights in these two cases, respectively. In both cases, once COSMIC data was assimilated with 300 km of the cyclone center track error started to decrease. Overall, COSMIC assimilation

runs (5 out of 8 cases) favored a leftward track bias (100 - 200 km) and lagged (50-200 km) GMA. In comparison, WCL-180hr and WR-180hr were more accurate and precise than WC-180hr, yet they too exhibited an average leftward track bias in 19 out of 32 (59.3%) simulations, but had smaller track error (50-150 km). Impressively large track errors in the Feb. 2009 (case 4) nor'easter were reduced by up to 1,400 and 1,700 km in the WCL-180hr and WR-180hr simulations, respectively, and were attributed to more frequently assimilated observations over the Southeastern US.

Simulated cyclone intensity was strongly linked to mid-troposphere latent heat release and its accuracy varied amongst assimilation runs. Between COSMIC cases, all but WC-72hr accurately simulated minimum SLP to within 5 hPa of GMA in 4 out of 8 cases, however longer COSMIC assimilation favored cyclone over-intensification. Comparisons between W32 and WC-180hr runs revealed the strength of simulated mid-tropospheric (~500 hPa) latent heat release was associated with changes in SLP tendencies. Strength of other factors (i.e., positive vorticity advection, jet quadrants, etc.) was not as frequently associated with SLP variations suggesting a less essential role. Cyclone over-intensification occurred in 2 out of 8 cases in WCL-180hr and for no cases in WR-180hr simulations. Minimum SLP in WCL-180hr and WR-180hr varied by upwards of 7.4 hPa (0.74%) and 22.7 hPa (2.3%), respectively, which was attributed to mid-tropospheric latent heat release variations. As compared to WC-180hr simulations, cyclone intensity was more accurate in 4 and 7 out of 8 cases in WCL-180hr and WR-180hr, respectively.

Storm- and GMA-centered energy norm differences proved an invaluable tool to evaluate nor'easter simulations accuracy. Assimilation of COSMIC improved up to 41

out of 61 (67.2%) WRF-centered periods, yet only 29 out of 61 (47.5%) GMA-centered periods. Energy norms based from WC-24hr did worst overall where at best it had less simulation error than W32 in 24 out of 61 (39.3%) periods. This result can be possibly associated with both residual model spin-up errors and in some cases a lack of observations near the cyclone center. Energy norm values typically were inversely proportional to assimilation period length due to both cyclone-to-sounding distance variations and sometimes small sample size. The largest energy norm contributions were around 850, 600, 300, and 180 hPa and were associated with errors in simulated planetary boundary layer height, intensity and positioning of the mid- and upper level jet and tropopause height. Direct comparison of WC-180hr to WCL-180hr and WR-180hr showed the latter had lower energy norms in 41 out of 60 (68.3%) and 48 out of 60 (80%) periods, respectively. Thus, the increased temporal error associated with the wider COSMIC window had no negative impact and simulations benefited more from radiosonde assimilation than from COSMIC alone due to the greater density of radiosonde observations.

The COSMIC assimilation energy norm analysis showed this quantity was typically inversely proportional to the period of data assimilation and not to the total number of observations. For WC-180hr simulations, case 6 exhibited the greatest mean energy norm reduction ($1.23 \times 10^{10} \text{ J m s}^{-2}$), yet it only had the fourth highest number of COSMIC soundings (755). In contrast, the case 8 had the most COSMIC observations (847), yet the 4th lowest mean reduction in energy norm ($1.00 \times 10^{10} \text{ J m s}^{-2}$). Overall, COSMIC simulations had a lower energy norm in as many as 93 (37.5%), 207 (83.5%), and 197 (79.4%) out of 248 periods in domains 1, 2, and 3, respectively. Consistent with

the local-storm environment results, WC-24hr performed worst and the major energy norm contributors were identical. Domain 1, however, experienced a greater proportion of its energy norm from the upper-level jet due its large spatial extent. In comparison to WC-180hr, WR-180hr and WCL-180hr had lower energy norms in at least 164 (66.3%) and 227 (91.5%) periods for domains 2 and 3, respectively. Unlike domains 2 and 3, domain 1 WCL-180hr energy norms were lower than WC-180hr in merely 66 (26.6%) periods owing to the assimilation of more poleward observations which afforded greater accumulation of stratospheric data assimilation errors. In stark contrast, WR-180hr also accumulated error from poleward locations on domain 1, but the spotty arctic station coverage mitigated its growth and resulted in it improving 223 (89.9%) periods as compared to WC-180hr.

With the exception of 2-PVU potential temperature on domains 2 and 3, RMSE results were largely consistent with the energy norm results. Domain 1 RMSE results however, indicated lower error in twice periods as compared to the energy norm results. Inconsistencies from 2-PVU potential temperature RMSE findings were attributed to stratospheric data assimilation errors. Combining RMSE (except 2-PVU potential temperature) and energy norm results, data assimilation for 48 hours or longer led to 50% or more periods in domains 2 and 3 having less error than W32. Due increased error associated with poleward observation assimilation, this same tipping point was 72 hours for domain 1. Except for SLP RMSE, WC-180hr was slightly better (11 additional improved periods [4.4%]) than WCL-180hr. Excluding 2-PVU potential temperature, WC-180hr simulations improved 35 (14.1%) fewer periods as compared to WR-180hr.

Overall, the above results have shown WRF nor'easter simulations can often benefit from COSMIC assimilation. Benefits to nor'easter simulation accuracy appear cumulative, require assimilation over at least 48 hours, and are particularly sensitive to sounding-to-cyclone distance, yet not directly associated to observation quantity. Therefore, WRF model simulations that assimilate data only at initialization within a regional domain gain little from COSMIC assimilation. More generally, WRF simulations that assimilate COSMIC observations from polar locations have the potential to be negatively impacted due to the lower stratospheric height, which augments stratospheric data assimilation error growth. Finally, although COSMIC is better able to obtain data from the local-storm environment, WRF simulations assimilating radiosondes had less error overall, which suggests COSMIC is not a full replacement for radiosondes. Despite this conclusion, it is worth noting that energy norm reductions from COSMIC runs were more than half that associated with radiosondes, which suggests that COSMIC is possibly more effective on a per observation basis than radiosondes for nor'easter simulations.

4.5. Future work

This work broadly addressed how COSMIC observations influenced nor'easter simulations, but several new questions can be raised. One such topic is observation sensitivity. A single COSMIC observation in case 4 helped correct an erroneously prolonged cut-off 500-hPa geopotential height minimum, but without this observation what would have happened? Future work could investigate how sensitive nor'easter simulations are to the proximity of COSMIC observations in general. Other work would

involve denying WRF COSMIC-based temperature or dewpoint and show its sensitivity to each. For this work, only radiosonde assimilation was compared to COSMIC, which elicits the question of how COSMIC compares to other assimilated data (i.e., microwave profilers, radiances, etc.). Impact of COSMIC upon other wintertime cyclones such as polar lows and in land winter storms (e.g., Colorado lows) could be evaluated and compared to nor'easters. Finally, this same analysis could also be completed using more advanced assimilation techniques (i.e., four-dimensional data assimilation or ensemble Kalman filters). From these runs, an investigation into how model improvement compared to additional computation cost could be conducted.

Table 10: Nor'easter case list. Column 3 shows the dates which each event impacted the northeastern US, while the last two columns denote the first and last times for each model run.

Case Number	NESIS	Event Dates	Model Run Start Date	Model Run End Date
1	N/A	22-24 Nov. 2006	11/19 12UTC	11/27 00UTC
2	2.54	15-17 Mar. 2007	3/12 18UTC	3/20 06UTC
3	N/A	15-17 Apr. 2007	4/12 06UTC	4/19 18UTC
4	1.65	1-2 Mar. 2009	2/26 12UTC	3/6 00UTC
5	N/A	15-16 Oct 2009	10/12 12UTC	10/20 00UTC
6	4.03	19-20 Dec. 2009	12/16 06UTC	12/23 18UTC
7	4.38	4-7 Feb. 2010	2/2 18UTC	2/10 06UTC
8	N/A	12-14 Mar. 2010	3/9 06UTC	3/16 18UTC

Table 11: GMA-relative storm track bias analysis. Values indicate the total the number six-hour time periods from all cases where each COSMIC model exhibited the shown track bias relative to GMA at each time period.

	Left		Right		Key	
Lead	16	26	17	18		
	22	22	24	20	W32	WC-0hr
	25	26	18	18	WC-24hr	WC-48hr
Lag	33	28	17	19	WC-72hr	WC-180hr
	28	25	20	19		
	19	22	27	27		

Table 12: Performance analysis relative to W32. Values denote the number and percentage of 6-hour periods where the energy norm or RMSE from the indicated simulations was lower than W32. The last section indicates the number of periods where COSMIC-based energy norms were lower than W32 within a 600 km box centered on GMA and each simulated storm, respectively. Bolded values indicate the best COSMIC simulation for each analysis component.

Domain 1 Performance Analysis vs WRF 3.2 (248 Total Times)	WC-0hr	WC-24hr	WC-48hr	WC-72hr	WC-180hr
Energy Norm	40 (16.1%)	39 (15.7%)	53 (21.4%)	75 (30.2%)	93 (37.5%)
RMSE 2 PVU Pot T	52 (21.0%)	53 (21.4%)	46 (18.5%)	36 (14.5%)	31 (12.5%)
RMSE 300-hPa Winds	68 (27.4%)	95 (38.3%)	121 (48.8%)	155 (62.5%)	205 (82.7%)
RMSE 500-hPa Geo Hght	75 (30.2%)	73 (29.4%)	105 (42.3%)	144 (58.1%)	200 (80.6%)
RMSE 850-hPa Temps	80 (32.3%)	112 (45.2%)	137 (55.2%)	155 (62.5%)	213 (85.9%)
RMSE SLP	97 (39.1%)	79 (31.9%)	112 (45.2%)	124 (50.0%)	160 (64.5%)
Domain 2 Performance Analysis vs WRF 3.2 (248 Total Times)	WC-0hr	WC-24hr	WC-48hr	WC-72hr	WC-180hr
Energy Norm	130 (52.4%)	107 (43.2%)	144 (58.1%)	168 (67.7%)	207 (83.5%)
RMSE 2 PVU Pot T	9 (3.6%)	17 (6.9%)	32 (12.9%)	25 (10.1%)	61 (24.6%)
RMSE 300-hPa Winds	90 (36.3%)	116 (46.8%)	149 (60.1%)	168 (67.7%)	199 (80.2%)
RMSE 500-hPa Geo Hght	78 (31.5%)	106 (43.7%)	142 (57.3%)	175 (70.6%)	204 (82.3%)
RMSE 850-hPa Temps	84 (33.9%)	127 (51.2%)	151 (60.9%)	184 (74.2%)	217 (87.5%)
RMSE SLP	103 (41.5%)	117 (47.2%)	147 (59.3%)	172 (69.4%)	183 (73.8%)
Domain 3 Performance Analysis vs WRF 3.2 (248 Total Times)	WC-0hr	WC-24hr	WC-48hr	WC-72hr	WC-180hr
Energy Norm	122 (49.2%)	102 (41.1%)	143 (57.7%)	171 (59.3%)	197 (79.4%)
RMSE 2 PVU Pot T	11 (4.4%)	14 (5.6%)	12 (4.8%)	15 (6.0%)	38 (15.3%)
RMSE 300-hPa Winds	117 (47.2%)	123 (49.6%)	147 (59.3%)	166 (66.9%)	188 (75.8%)
RMSE 500-hPa Geo Hght	102 (41.1%)	109 (44.0%)	137 (55.2%)	170 (68.5%)	191 (77.0%)
RMSE 850-hPa Temps	84 (33.9%)	110 (44.4%)	136 (54.8%)	168 (67.7%)	201 (81.1%)
RMSE SLP	111 (44.8%)	123 (49.6%)	132 (53.2%)	159 (64.1%)	164 (66.1%)
Domain 2 Storm Performance Analysis vs WRF 3.2 (61 Total Times)	WC-0hr	WC-24hr	WC-48hr	WC-72hr	WC-180hr
Energy Norm GMA Rel	29 (47.5%)	24 (39.3%)	29 (47.5%)	27 (44.3%)	29 (47.5%)
Energy Norm WRF Rel	32 (52.5%)	23 (37.7%)	37 (60.7%)	37 (60.7%)	41 (67.2%)

Table 13: Same as Table 11, except it includes only the 180-hr model runs and W32.

	Left		Right		Key	
Lead	16	25	17	19		
	22	39	24	14	W32	WC-180hr
Lag	33	20	17	28	WCL-180hr	WR-180hr
	17	20	30	14		

Table 14: Same as Table 3, except it includes only full assimilation simulations. Unlike Table 12, this analysis included runs with a ± 1.5 hour COSMIC assimilation window and radiosonde assimilation runs (noted as RAOB). The ± 1 hour window COSMIC assimilation is the same simulation as used in Table 12. Bolded values indicate the best performing data assimilation run for each analysis component.

Domain 1 Performance Analysis vs WRF 3.2 (248 Total Times)	WC-180hr	WCL-180hr	WR-180hr
Energy Norm	93 (37.5%)	82 (33.1%)	184 (74.2%)
RMSE 2 PVU Pot T	31 (12.5%)	32 (12.9%)	61 (24.6%)
RMSE 300-hPa Winds	205 (82.7%)	206 (83.1%)	211 (85.1%)
RMSE 500-hPa Geo Hght	200 (80.6%)	198 (79.8%)	214 (86.3%)
RMSE 850-hPa Temps	213 (85.9%)	202 (81.5%)	228 (91.9%)
RMSE SLP	160 (64.5%)	166 (66.9%)	200 (80.6%)
Domain 2 Performance Analysis vs WRF 3.2 (248 Total Times)	WC-180hr	WCL-180hr	WR-180hr
Energy Norm	207 (83.5%)	204 (82.3%)	231 (93.1%)
RMSE 2 PVU Pot T	61 (24.6%)	62 (25.0%)	81 (32.7%)
RMSE 300-hPa Winds	199 (80.2%)	201 (81.0%)	215 (86.7%)
RMSE 500-hPa Geo Hght	204 (82.3%)	201 (81.0%)	212 (85.5%)
RMSE 850-hPa Temps	217 (87.5%)	213 (85.9%)	228 (91.9%)
RMSE SLP	183 (73.8%)	180 (72.6%)	212 (85.5%)
Domain 3 Performance Analysis vs WRF 3.2 (248 Total Times)	WC-180hr	WCL-180hr	WR-180hr
Energy Norm	197 (79.4%)	200 (80.6%)	230 (92.7%)
RMSE 2 PVU Pot T	38 (15.3%)	32 (12.9%)	35 (14.1%)
RMSE 300-hPa Winds	188 (75.8%)	187 (75.4%)	217 (87.5%)
RMSE 500-hPa Geo Hght	191 (77.0%)	186 (75.0%)	206 (83.1%)
RMSE 850-hPa Temps	201 (81.1%)	194 (78.2%)	221 (89.1%)
RMSE SLP	164 (66.1%)	175 (70.6%)	199 (80.2%)
Domain 2 Storm Performance Analysis vs WRF 3.2 (60 Total Times)	WC-180hr	WCL-180hr	WR-180hr
Energy Norm GMA Rel	35 (58.3%)	43 (71.7%)	48 (80.0%)
Energy Norm WRF Rel	40 (66.7%)	43 (71.7%)	53 (88.3%)

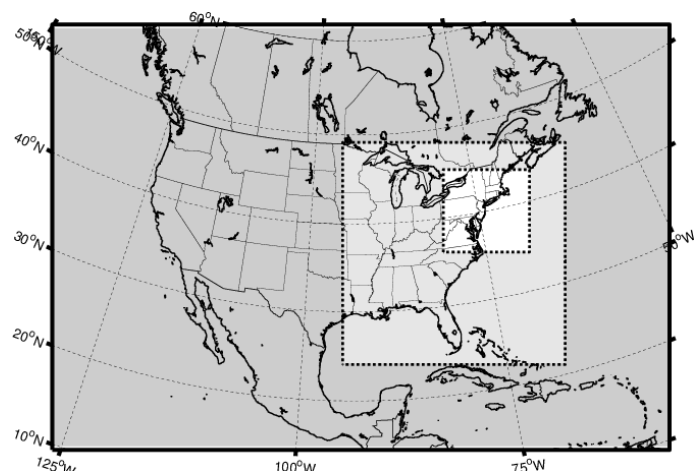


Fig. 31: Nested grid configuration used in simulations. Horizontal grid spacing for domains 1, 2, and 3 are 45, 15, and 5 km, respectively.

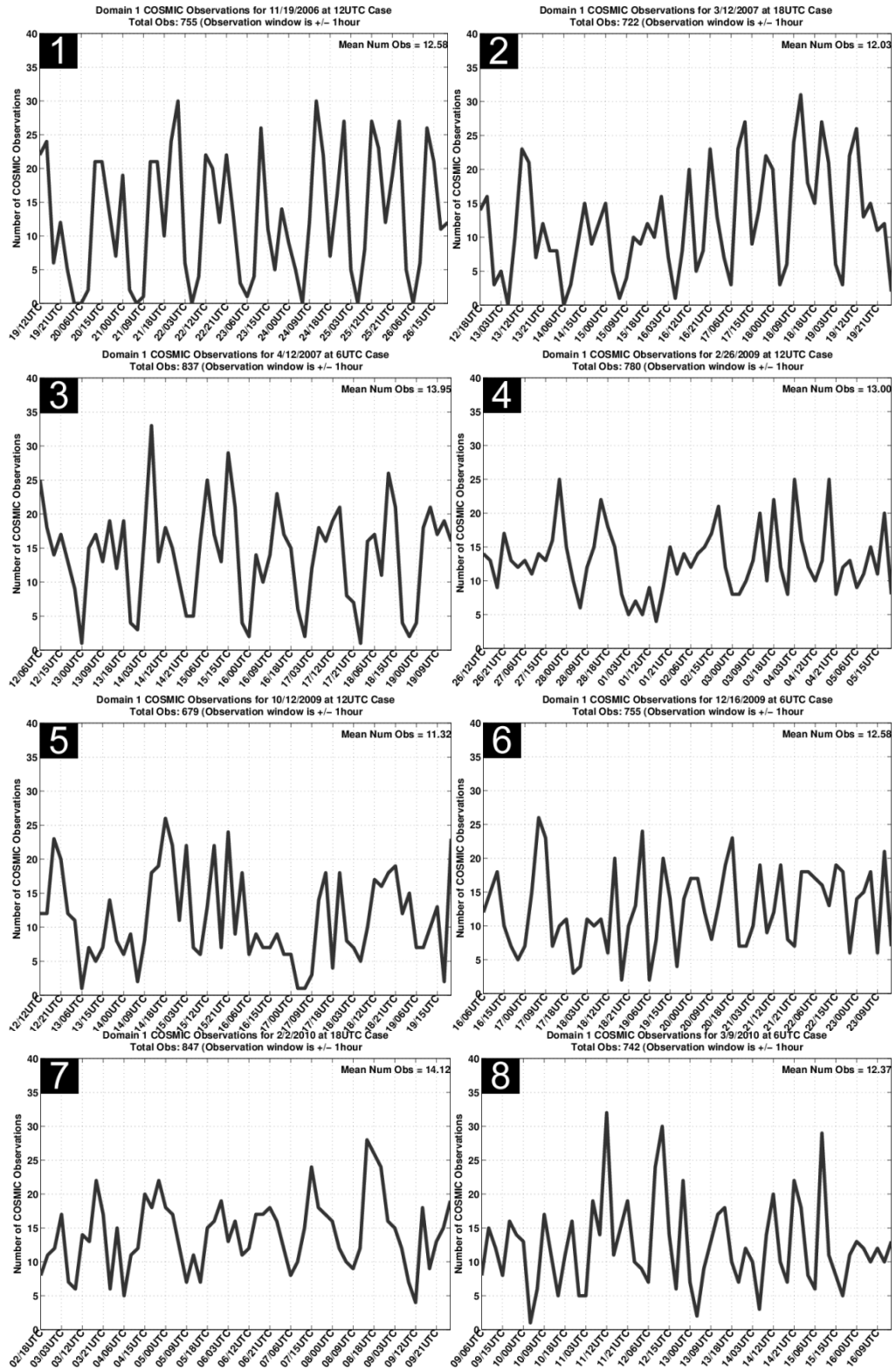


Fig. 32: Number of COSMIC data assimilations within a ± 1 hour window every three model hours on domain 1. White number indicate case number.

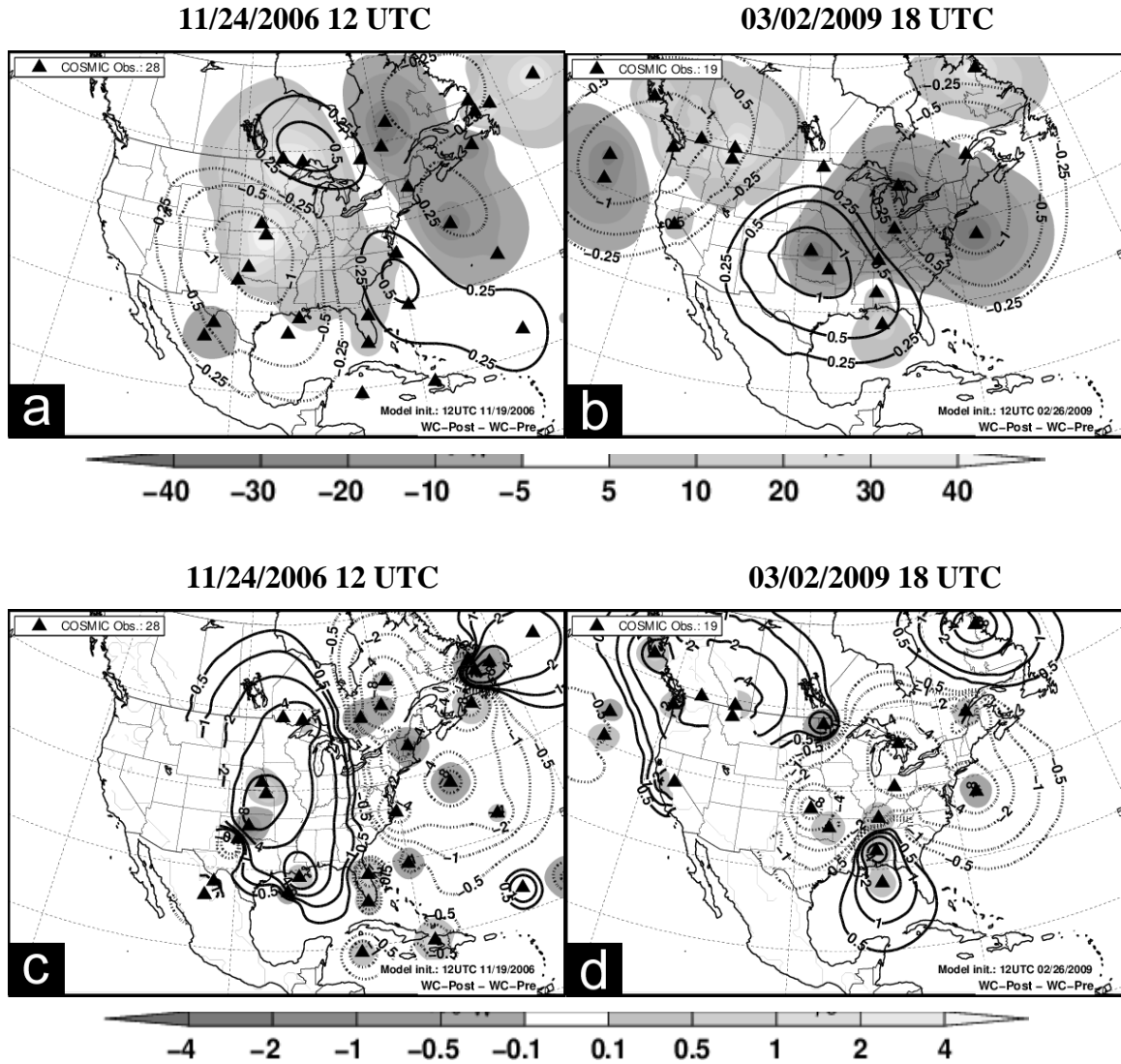


Fig. 33: COSMIC analysis increments from (left) 12 UTC 24 November and (right) 18 UTC 2 March 2009. (top) Differences in 500-hPa geopotential height (fills, m) and sea-level pressure (contours, hPa). (bottom) Differences in 850-hPa water vapor (fills, g kg⁻¹) and temperature (contours, K). Triangles denote COSMIC profile locations.

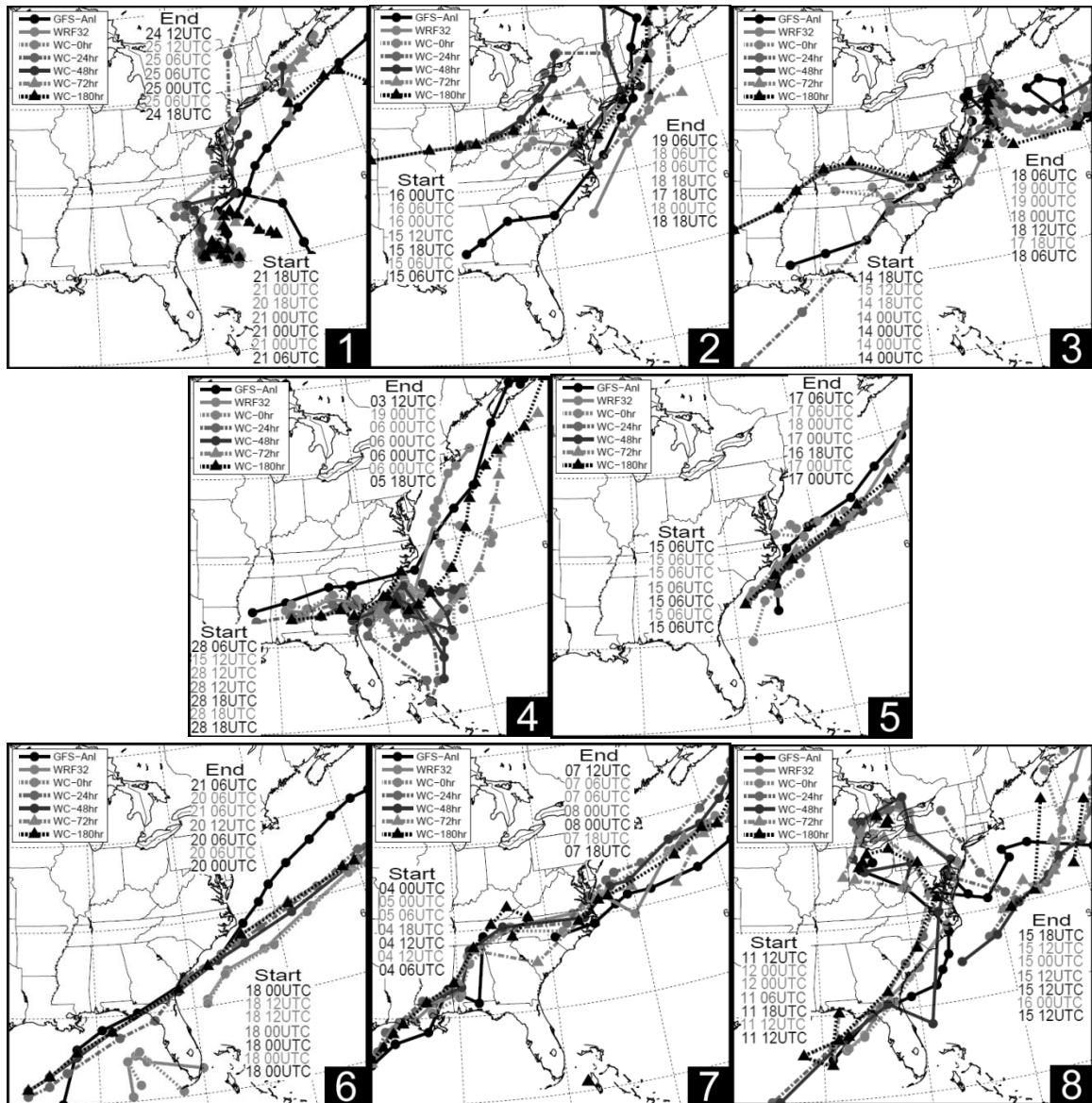


Fig. 34: Storm tracks from GMA and each COMSIC assimilation run. Line legend is shown on the upper-left of each plot. Shown symbols indicate simulated storm position every six hours. White numbers indicate case number.

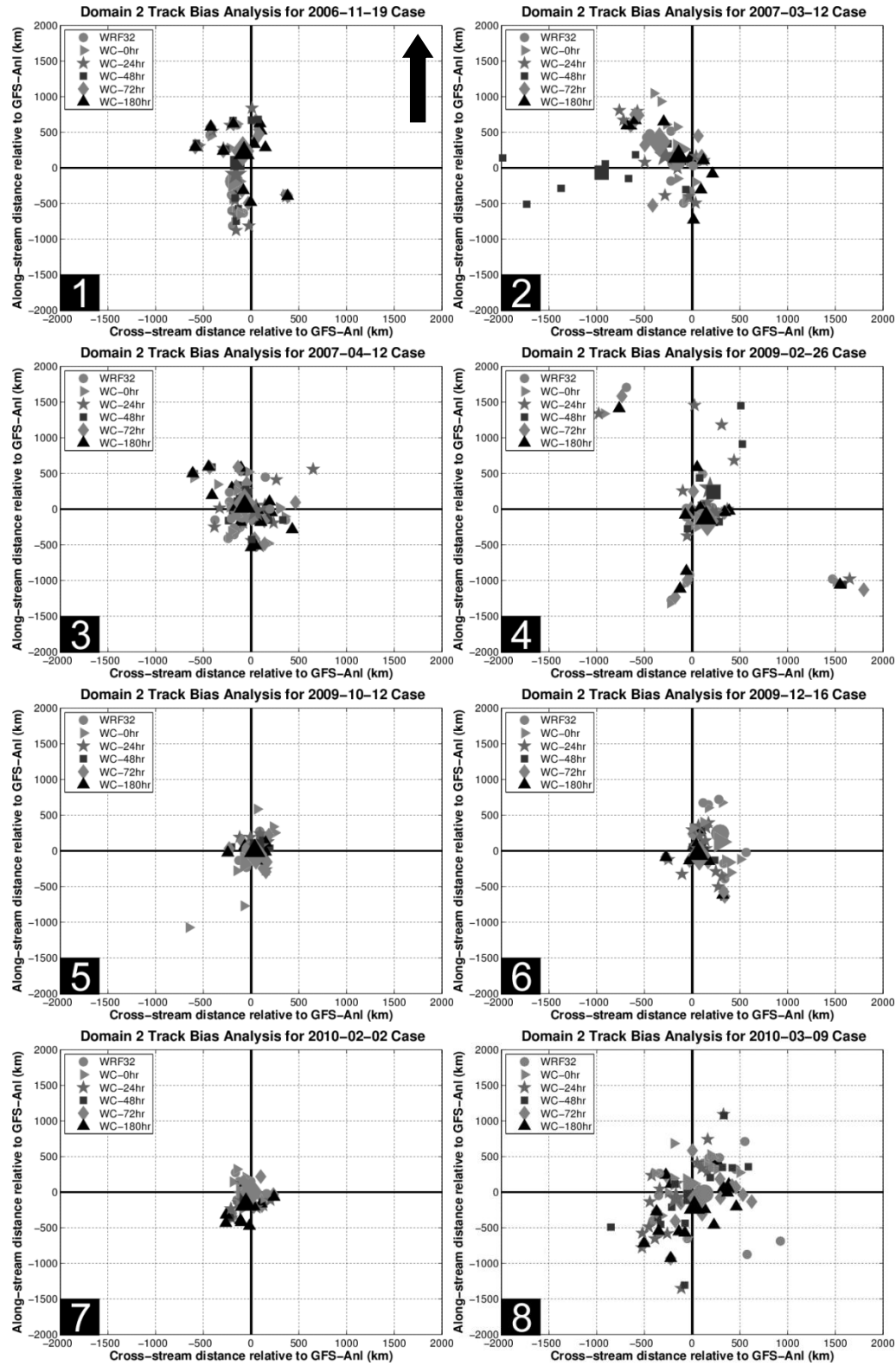


Fig. 35: WRF forecasted storm position bias as compared to GMA for all eight cases. The black arrow in panel 1 indicates the GMA storm motion direction for all panels. Shown symbols represent WRF position bias every six hours (smaller symbols) and their mean (large symbols).

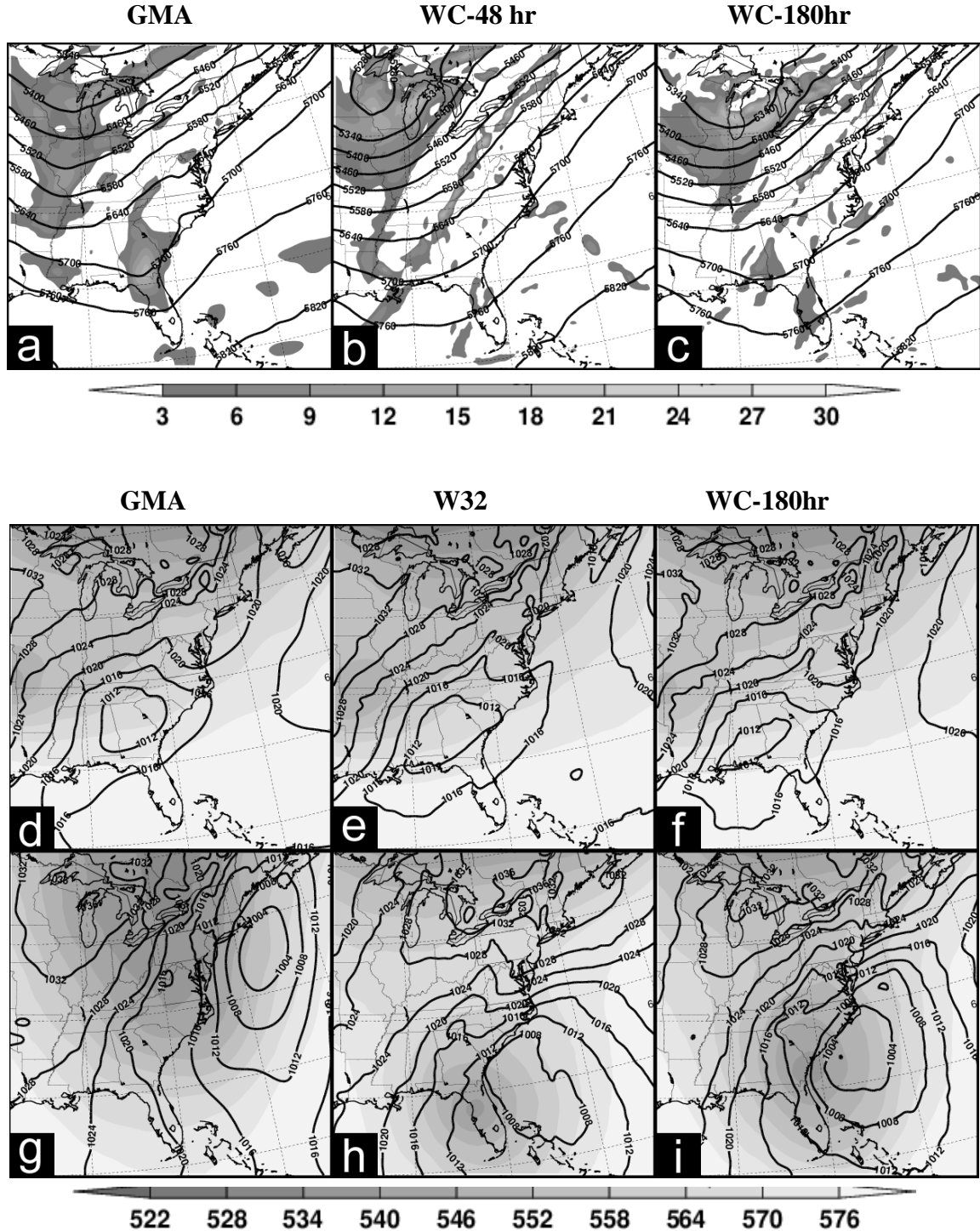


Fig. 36: Various plots from (a,d,g) GMA, (e,h) W32, (b) WC-48 hr (B), and (c,f,i) WC-180hr. (a-c) 500-hPa relative vorticity (shaded, s^{-1}) and geopotential height (contours, m) on 12 UTC 16 March 2007. (d-f) 500-hPa geopotential height (shaded, dam), and sea-level pressure (hPa, contours) on 18UTC 28 Feb. 2009. (g-i) same as (d-f) except for 18 UTC 2 March 2009 at 18UTC.

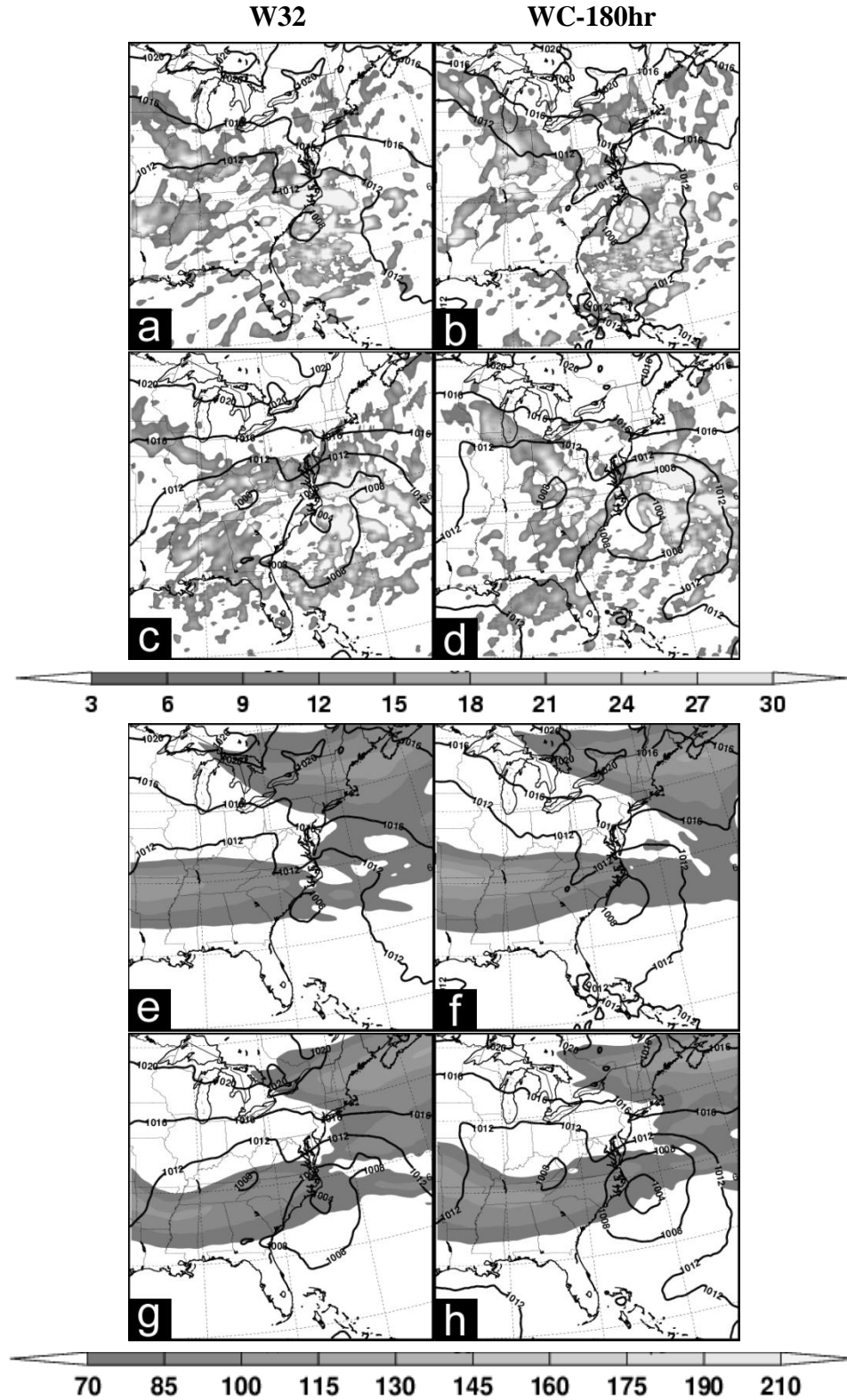


Fig. 37: (a-d) 500-hPa latent heating (fills, K day⁻¹) and sea-level pressure (contours, hPa) and (e-h) 300-hPa winds (fills, knots) and sea-level pressure (contours, knots) (E-H). (a,b,e,f) are from 12 UTC 15 October 2009 and (c,d,g,h) are from 18 UTC the same day.

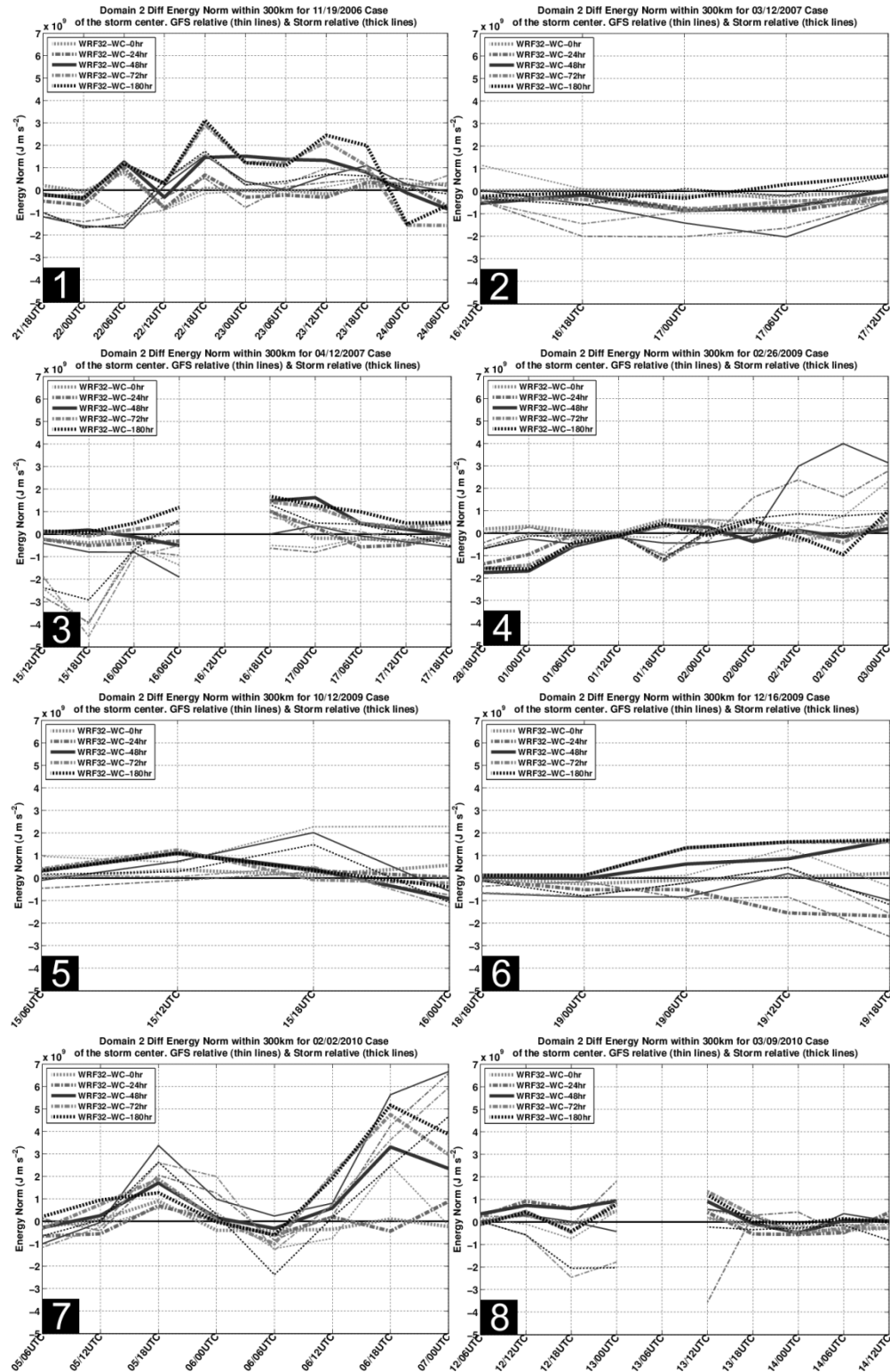


Fig. 38: Domain 2 differences in COSMIC run energy norms for all cases within 300 km of the GFS analysis storm center (thin lines) and each model simulated center (thick lines). Shown differences are relative to WRF 3.2. Positive values denote improvement over WRF 3.2.

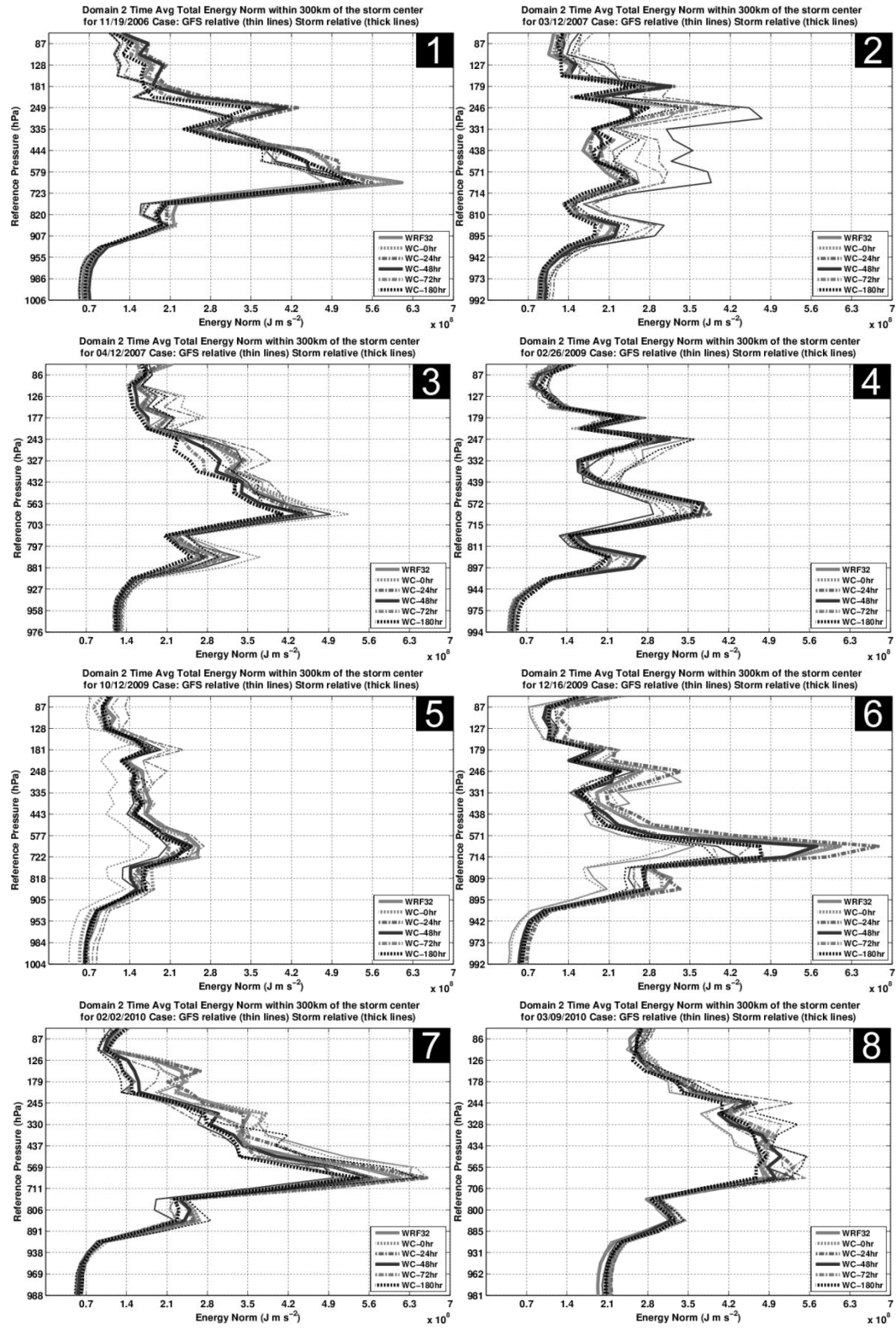


Fig. 39: Domain 2, time-averaged energy norm for all cases with 300 km of the GFS analysis storm center (thin lines) and each model simulated center (thick lines).

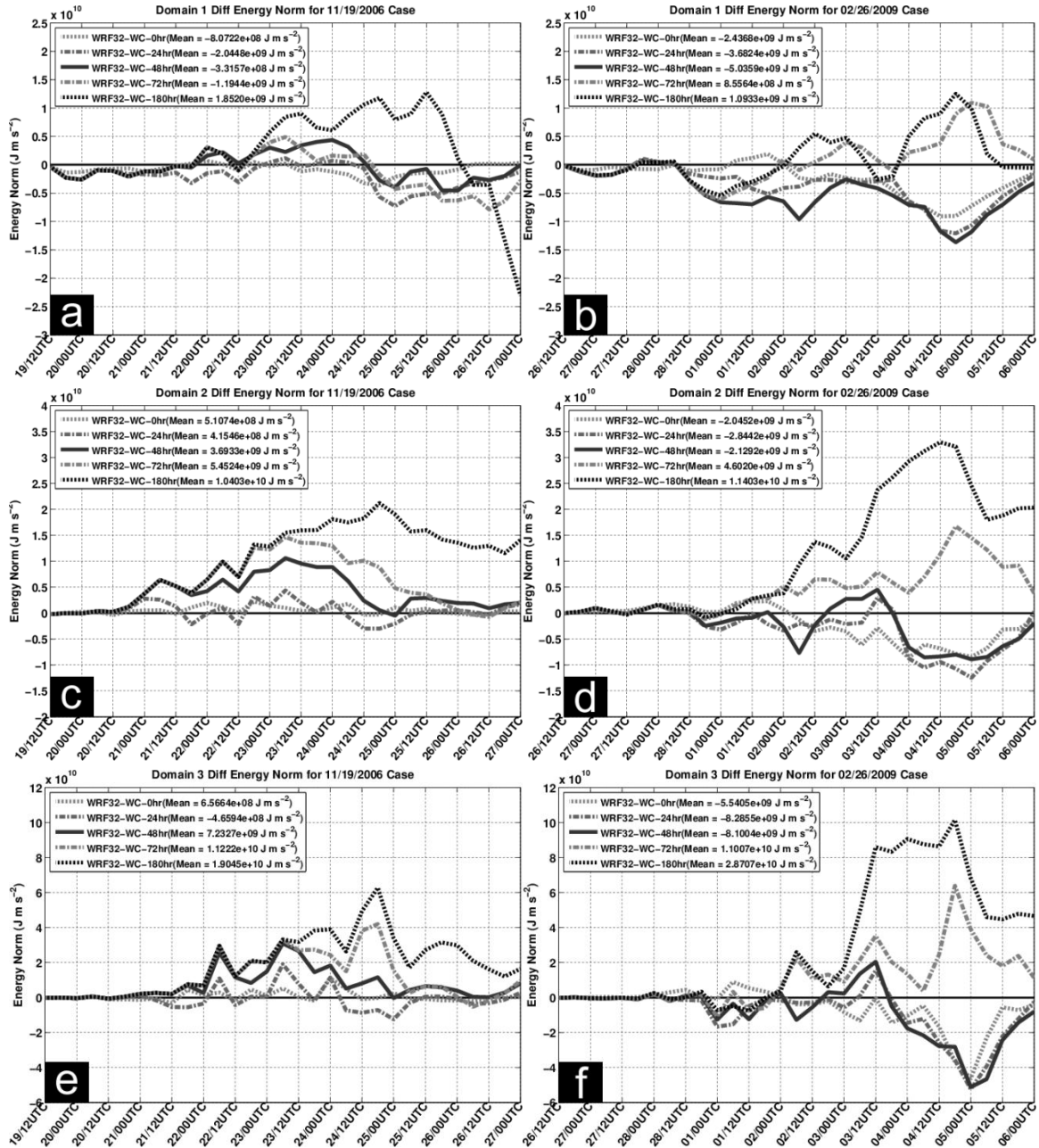


Fig. 40: Energy norm differences in COSMIC runs within the entirety of domains 1, 2, and 3 for (a, c, e) case 1 and (b, d, e) case 4. Shown differences are relative to W32. Positive values denote improvement over W32.

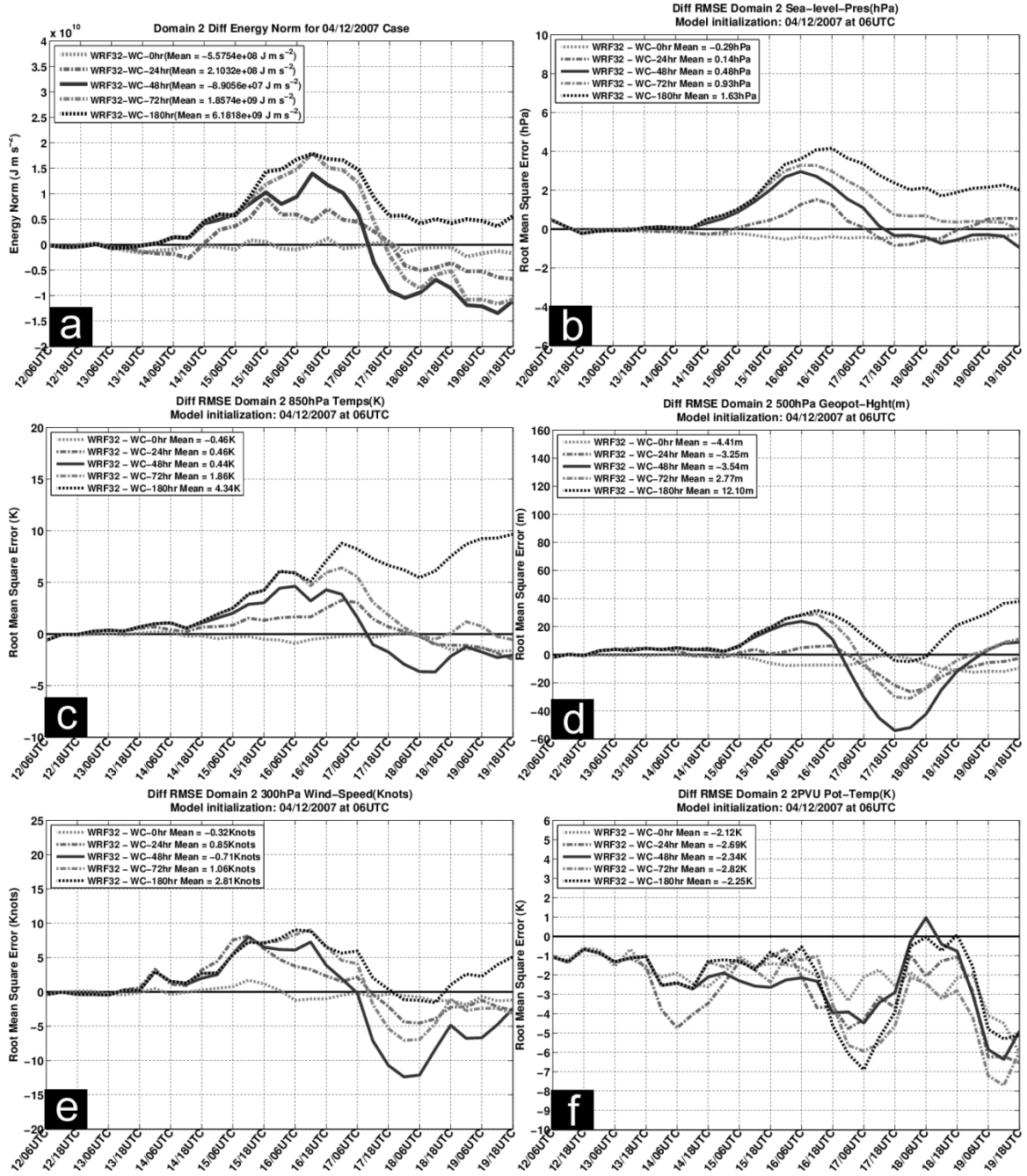


Fig. 41: Difference in domain 2 (a) energy norm, (b) sea-level pressure, (c) 850-hPa temperature, (d) 500-geopotential height, (e) 300-hPa winds, and (f) 2PVU potential temperature for case 3. Shown differences are relative to W32. Positive values denote improvement over W32.

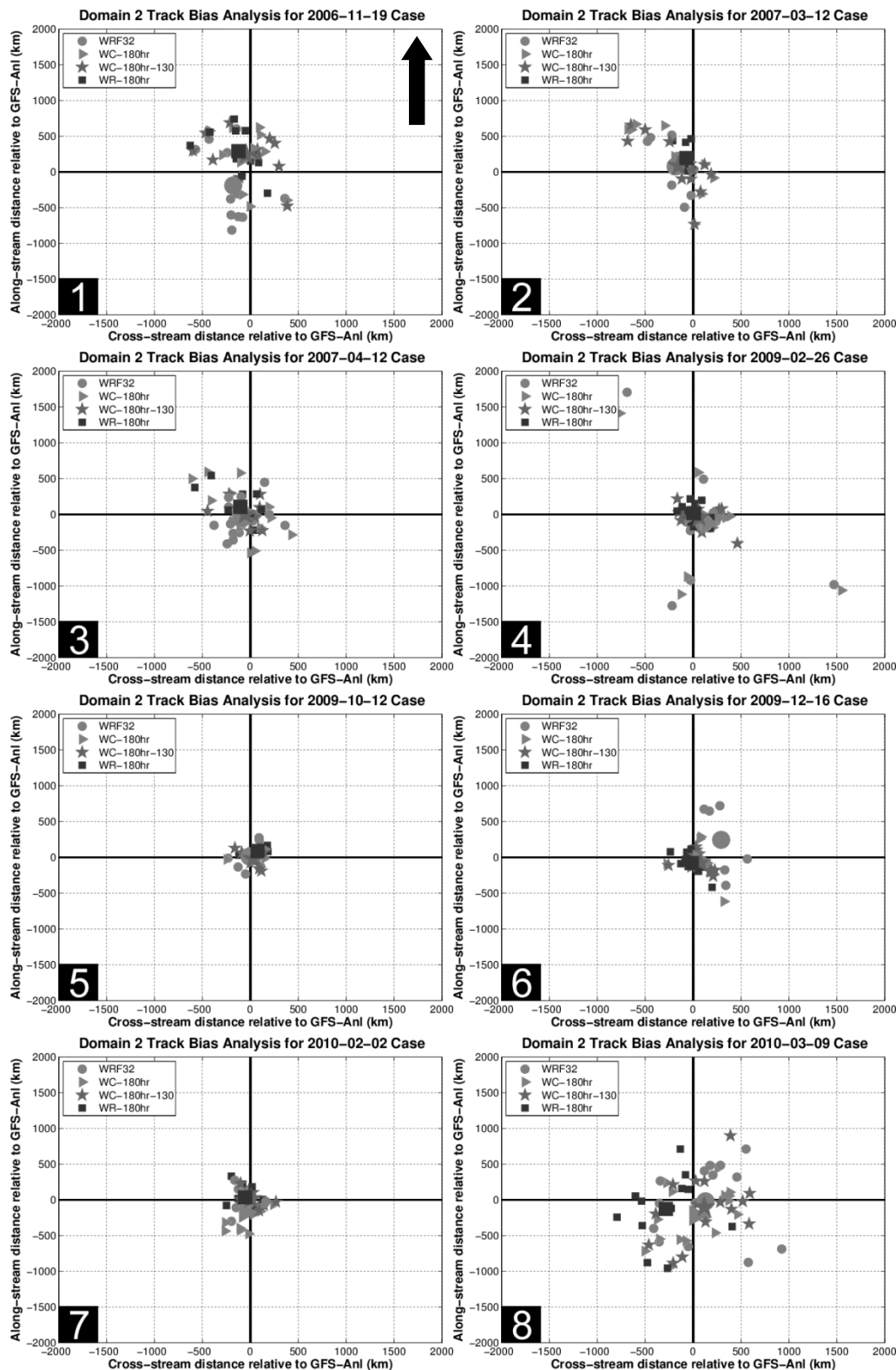


Fig. 42: WRF forecasted storm position bias as compared to GMA for all eight cases. The black arrow in panel 1 indicates the GMA storm motion direction for all panels. Shown symbols represent WRF position bias every six hours (smaller symbols) and their mean (large symbols).

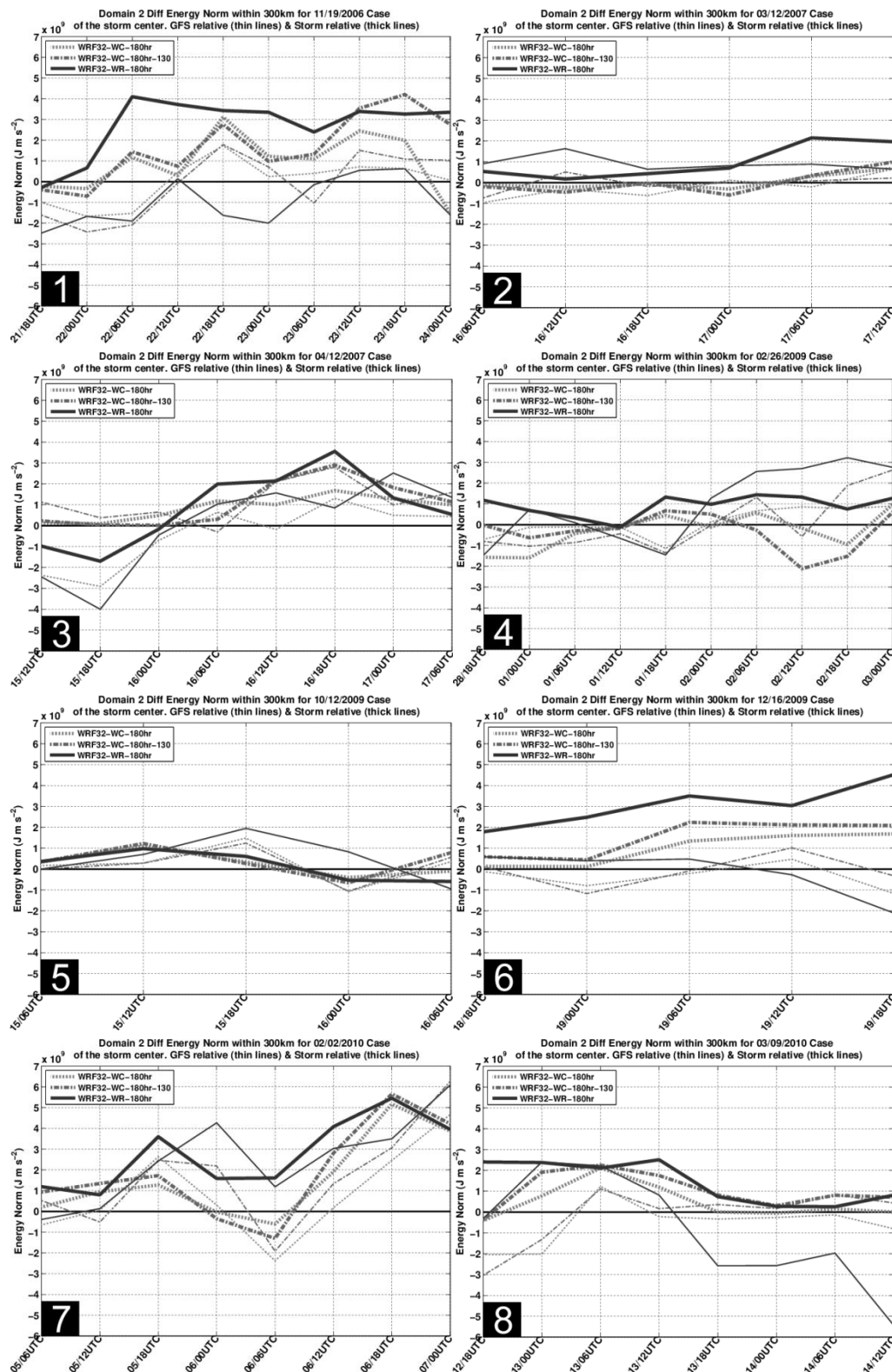


Fig. 43: Domain 2 differences in 180 hour COSMIC (1 and 1.5 hour window) and radiosonde assimilation energy norms for all cases with 300 km of the GFS analysis storm center (thin lines) and each model simulated center (thick lines). Shown differences are relative to WRF 3.2. Positive values denote improvement over WRF 3.2.

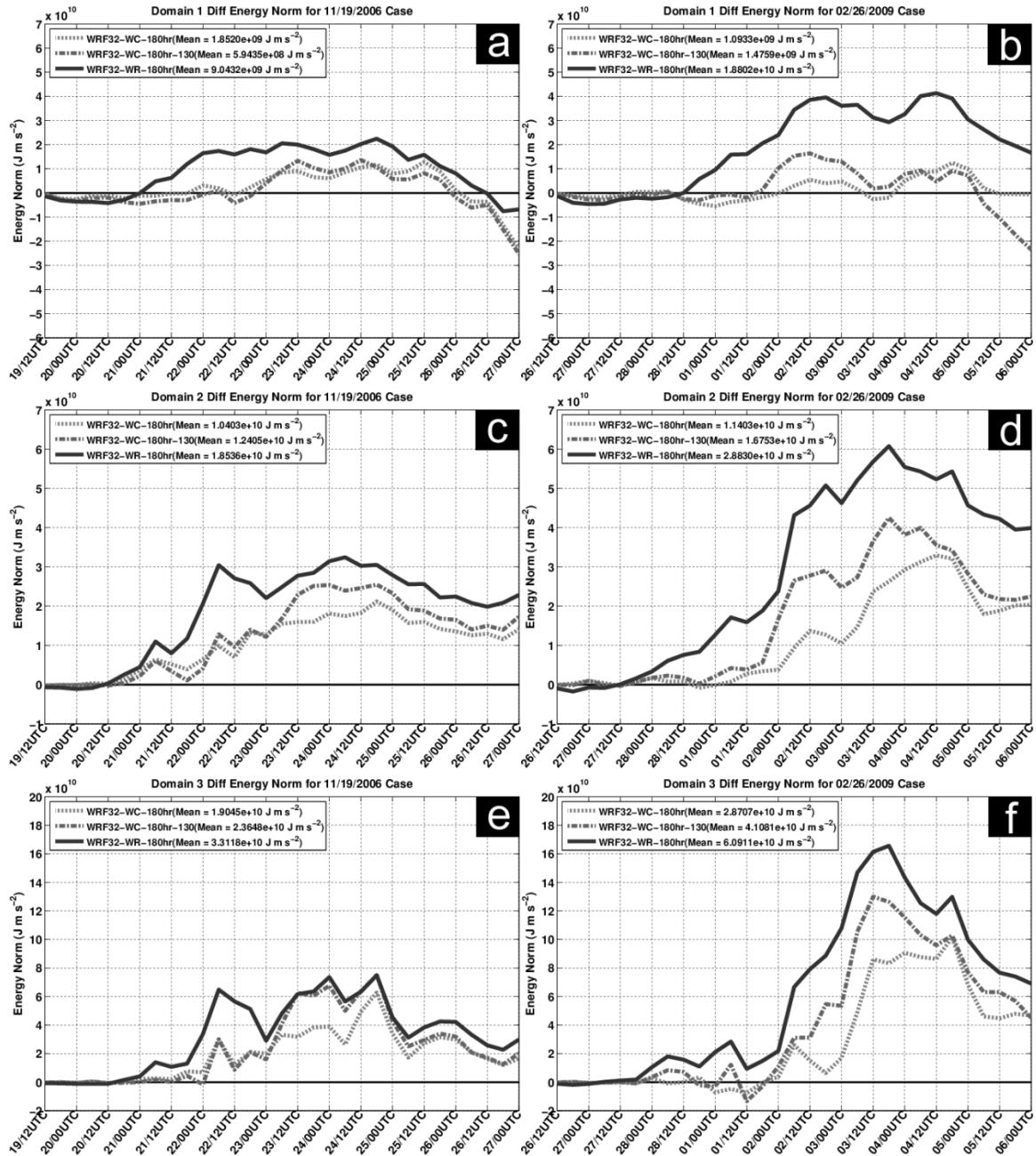


Fig. 44: Energy norm differences in 180 hour COSMIC (1 and 1.5 hour window) and radiosonde assimilation energy norms COSMIC runs within domains 1, 2, and 3 for case 1 (a,c,e) and case 4 (b,d,e). Shown differences are relative to W32. Positive values denote improvement over W32.

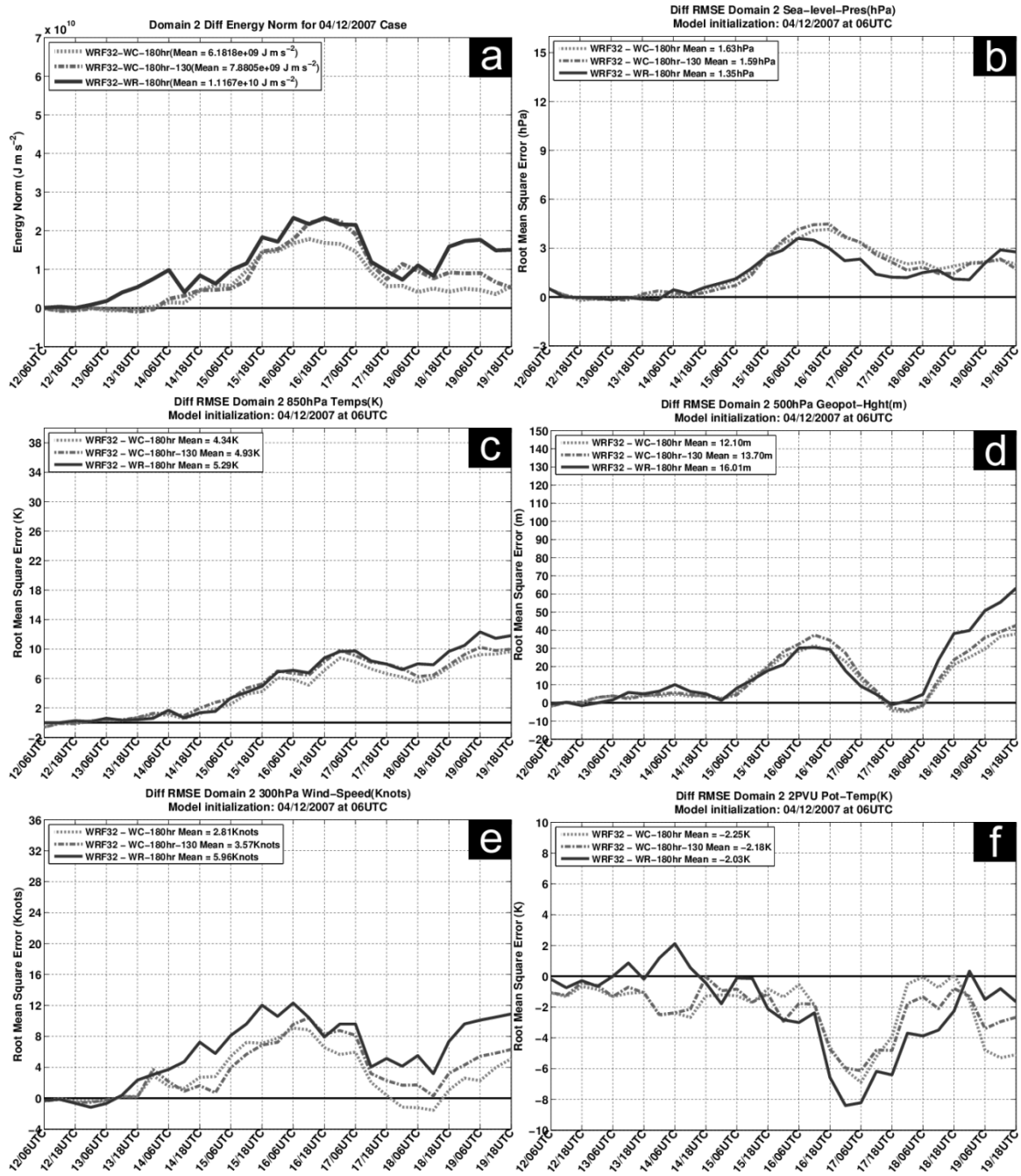


Fig. 45: Difference in domain 2, 180 hour COSMIC (1 and 1.5 hour window) and radiosonde assimilation (a) energy norm, (b) sea-level pressure, (c) 850-hPa temperature, (d) 500-geopotential height, (e) 300-hPa wind, and (f) 2PVU potential temperature for case 3. Shown differences are relative to W32. Positive values denote improvement over W32.

5. Impact of coupling an ocean model to WRF nor'easter simulations

5.1 Chapter abstract

An investigation of how ocean-atmosphere coupling impacted Weather Research and Forecasting (WRF) Model simulations of intense, winter time cyclones was completed. Simulations were conducted for 180 hours starting roughly 72 hours prior to the first precipitation impacts in the highly populated Mid-Atlantic US and associated cyclogenesis and verified against GFS model analysis. Model simulations included an uncoupled WRF run, a coupled 1D mixed layer, and a coupled, full dynamics, 3D ocean model simulation. Over the entire 180 hour simulation the magnitude of sea-level pressure, 10 m winds, and 500-hPa geopotential height varied from uncoupled model simulations upwards of 6.0 hPa, 14.5 m s^{-1} , and 45 m, respectively. A majority of WRF simulations (12 out of 18) favored a leftward bias and (10 out of 12) lagged Global Forecasting System Model Analysis, but no notable storm track changes were observed. Coupling WRF to a full physics ocean model however, did lead to slightly improved simulations of minimum sea-level pressure associated primarily with adjustments to mid-tropospheric latent heat release. Although ocean-atmosphere coupling improved up to 54.3% of six-hour periods over uncoupled WRF simulations, differences in energy norms, RMSEs, and various atmospheric fields indicated only subtle changes in both simulation outcome and accuracy. Finally, alterations to dynamical fields due to coupling appears limited to regions of the atmosphere at and below 500 hPa.

5.2. Methods

5.2.1. COAWST Overview

For this study we utilized COAWST Version 3 (Warner et al. 2010). As shown in Fig. 46, COAWST currently has four WRF configuration options (1 uncoupled, 3 coupled). Of these four options, we used only the WRF only and WRF-ROMS coupling configurations (Figs. 46a-b). As the focus of this work is solely on ocean-atmosphere coupling and its impact on nor'easter simulations, we elected neither use either SWAN or CSTMS, but may do so in future work.

Because the WRF and ROMS grids (Fig. 47) are not co-located, the Spherical Coordinate Remapping Interpolation Package (SCRIP; Jones 1998) was used to compute interpolation weights for WRF and ROMS, making coupling via the MCT possible. Specific details about the WRF and ROMS configurations are explained below in sections b and c, respectively.

5.2.2. WRF configuration

As the COAWST atmospheric component, we utilized WRF Advanced Research WRF version 3.2.1 (hereafter W321), which numerically solves a set of fully-compressible, non-hydrostatic, Eulerian equations in terrain-following coordinates on an Arakawa C-Grid (Skamarock et al. 2008; Wang et al. 2012). To simulate a multi-scale phenomenon such as a nor'easter, model grids of sufficient size and resolution were required. The three-domain grid (Fig. 47a) had two-way interaction, 45, 15, and 5 km grid spacing, respectively, 27 vertical levels, and a 50-hPa (~20 km) model top. This configuration afforded simulation of key pre-cursor synoptic and meso- α -scale

phenomena (e.g., jet streaks, short- and long-wave troughs) on outer domains, and smaller-scale phenomena (e.g., orographic forcing, latent heating) on inner domains. Atmospheric boundary conditions were derived from $1^\circ \times 1^\circ$ Global Forecast System (GFS) model forecasts (<ftp://nomads.ncdc.noaa.gov/GFS/Grid3>). Ocean boundary data were RTG-SST applied along the WRF bottom boundary and included only SST data. For WRF-ROMS coupled simulations, ROMS-based SSTs were applied to the coupled WRF domain (domain 2) only where it overlapped the ROMS model grid (Fig. 47b). For regions of domain 2 outside the ROMS grid and on domains 1 and 3, RTG-SST was utilized. We chose to couple ROMS to WRF domain 2 because COAWST is unable to couple ROMS to multiple WRF domains and domain 2 was the only model grid of sufficient grid spacing (15 km), which also simulated the entire US East Coast and more importantly the Gulf Stream. Time step size for all three WRF model domains was 90, 30, and 10 seconds, respectively to match the time step size of the coupled ROMS domain and to be consistent with Warner et al. (2010).

Model parameterizations were consistent with past and on-going WRF studies at GSFC (e.g., Shi et al. 2010) with one deviation: The shortwave radiation scheme. On-going work at GSFC (i.e., Shi et al. 2010) uses the Goddard longwave scheme (Chou and Suarez 1999), but this scheme was not implemented until WRF version 3.3, instead we utilized the Dudhia shortwave scheme (Dudhia 1989). The WRF model options used in this study include:

- Longwave radiation: New Goddard Scheme (Chou and Suarez 1999; Shi et al. 2010)
- Shortwave radiation: Dudhia (Dudhia 1989)

- 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: Grell-Devenyi ensemble scheme (Grell and Devenyi 2002)
- Microphysics: Goddard (Lang et al. 2011)

5.2.3. ROMS configuration

As the COAWST ocean component, we utilized ROMS release 455 (Shchepetkin and McWilliams 2005; Haidvogel et al. 2008), which numerically integrates three-dimensional Reynolds-averaged Navier–Stokes equations using hydrostatic and Boussinesq approximations along vertically stretched, terrain-following coordinates on a horizontal curvilinear Arakawa C-grid. Given the importance of heat fluxes and momentum exchanges to nor’easter simulations, the ROMS model domain needed wide spatial coverage, sufficient grid spacing, and remain within the bounds of WRF domain 2 (Fig. 47a) for stability. The ROMS domain (Fig. 47b) has 8-km grid spacing, 16 sigma levels, a 5 m sigma-coordinate surface control parameter, a 0.4 m sigma-coordinate bottom control parameter and a 50 m sigma-coordinate surface/bottom layer width. For consistency with Warner et al. (2010), ROMS utilized a 30 second baroclinic time, and a time splitting ratio of 30m, and utilized open-boundary conditions.

ROMS initialization and boundary conditions required three primary data inputs: Atmospheric state variables, oceanic state variables, and tidal information. For WRF-ROMS coupling simulations, atmospheric state variables were passed directly from WRF

to ROMS and included wind, pressure, moisture, radiation, and temperature. Ocean initialization and boundary condition information were derived from the global Hybrid Coordinate Ocean Model (HYCOM; <http://hycom.coaps.fsu.edu/thredds/catalog.html>). Data retrieved from HYCOM included: Currents, water level, salinity, and temperature. Tidal initialization information was obtained from the Advanced Circulation (ADCIRC; Mukai et al. 2001) tidal data base which provided the five tidal constituents (M2, N2, S2, O1, and K1) required for open boundary conditions. Our choice of HYCOM and ADCIRC to initialize ROMS was for consistency with Warner et al. (2010).

5.2.4. Cases for analysis, model run description

As is Nicholls and Decker (2012), this study focused on the eight nor'easter cases seen in Table 15, but due to missing HYCOM data cases 2 and 3 were excluded from the analysis. Case selection was based upon three main factors. All cases occurred after an April 2006 satellite launch to correlate with a parallel nor'easter study (Nicholls and Decker 2012) and during the “nor'easter season” (October to April) as defined by Jacobs et al. (2005). To make this study more generalized, we did not select only the most severe nor'easters, but instead aimed for a relatively small but diverse sample. Case severity was measured using the Northeast Snowfall Impact Scale (NESIS; Kocin and Uccellini 2004). Under NESIS, nor'easters were classified on a scale of 1 (notable) to 5 (extreme) based upon the population impacted, area affected, and event severity. Finally, all cases were required to have precipitated in New Jersey.

Model simulations were conducted for 180 hours, starting roughly 72 hours prior to the first precipitation impacts in the highly populated Mid-Atlantic US and associated

cyclogenesis. This time frame focused attention exclusively on cyclone initiation and its later impact in this region. A 72-hour lead time allowed simulations to spin-up, establish baroclinicity between the cooler eastern United States and warmer Gulf Stream, and simulate latent heating along the expansive (>1000 km) northern edge of the Gulf Stream. All of the above are vital for accurate nor'easter simulations (Kuo et al. 1991; Mote et al. 1997; Yao et al. 2008). Precipitation data from the New Jersey Weather and Climate Network (Robinson 2005) served as a proxy for establishing when each nor'easter first impacted the Mid-Atlantic US. Using these data, model initialization was set as 72 hours prior to the first nor'easter-related 0.5 mm (~ 0.02 inch) precipitation reading. A New Jersey-centric approach was chosen because of its high population density ($461.6/\text{km}^2$), significant contribution (\$473 billion) to the US gross domestic product, and its relatively central location in the region of interest (United States Census Bureau 2012).

To assess the impact of ocean coupling on WRF simulations, three model runs were completed for the six nor'easter cases. These three runs included WRF-only (W321), coupled WRF-Pollard 1D mixed layer model (WPol; Pollard et al. 1972), and coupled WRF-ROMS (COAWST) simulations. Each of these simulations was configured identically except for the degree of ocean-atmosphere coupling. For W321, SST data was obtained from RTG-SST at model initialization and was fixed for the entire simulation. The WPol runs updated the mixed layer depth every time step and with SSTs being forced to match RTG-SST every time the boundary conditions were updated. Mixed layer depth for all WPol simulations was initialized at 200 m based upon a 150 year Climate Model Inter-comparison Model project average (October – April). To specifically diagnose how the Pollard model impacted WRF simulations, SST update was not active

for WPol simulations in this study. Finally, COAWST simulations had WRF and ROMS exchange data once every ten minutes for consistency with Warner et al. (2010).

5.2.5. Verification and analysis techniques

Atmospheric validation data was derived from GFS model analysis (GMA). We favored these data because all model domains included data-sparse regions where in-situ observations were often not available and GMA was easily interpolated to the WRF model grid. Furthermore, the atmospheric lateral boundary conditions were based on GFS forecasts. Because this study focuses solely upon the impact of ocean coupling on WRF model simulations, we decided to not validate ROMS model output in this study, but may do so in future research.

Model run analysis was comprised of several parts. Sea surface temperature data from the three model runs were compared to assess how each simulation differed and to determine how SSTs evolved during each simulation. Storm tracks were determined using local SLP minima via an objective, self-developed algorithm similar to that used at the Climate Prediction Center (Serreze 1995; Serreze et al. 1997). Quantifying storm track simulation differences included finding the minimum storm SLP and GMA-relative track bias. Overall accuracy of the local-storm environment (i.e., within a 600-km wide, WRF-centered box) and large-scale environments (i.e., entire model domain) were evaluated using the dry energy norm (Rabier et al. 1996).

$$\langle \mathbf{X}, \mathbf{Y} \rangle = -\frac{1}{2} \int_{p_{sfc}}^{p_{top}} \iint_A \left(\Delta u^2 + \Delta v^2 + R_d T \Delta \ln(p_{sfc})^2 + \frac{c_p}{T_v} \Delta T^2 \right) dx dy dp \quad (1)$$

where \mathbf{X} is the WRF model state vector, \mathbf{Y} is the GMA state vector, u is the zonal wind (m s^{-1}), v is the meridional wind (m s^{-1}), R_d is the dry air gas constant ($287 \text{ J kg}^{-1} \text{ K}^{-1}$)

¹), T_r is the mean surface temperature (K), P_{sfc} is the surface pressure (Pa), c_p is heat capacity at constant pressure ($1004 \text{ J kg}^{-1} \text{ K}^{-1}$), and T is air temperature (K). A 600-km wide box was used to evaluate the local-storm environment because it captured the storm, yet minimized background environment contamination. Root mean square error (RMSE) calculations were performed for SLP, 850-hPa temperature, 500-hPa geopotential height, 300-hPa winds, and 2 potential vorticity unit (2 PVU; $1 \text{ PVU} = 10^{-6} \text{ K m}^2 \text{ kg}^{-1} \text{ s}^{-1}$) potential temperature. The first four variables were frequently referenced for nor'easter analyses in Kocin and Uccellini (2004) and 2-PVU potential temperature was used to investigate changes to the dynamic tropopause. Despite their mathematical differences smaller values in both error metrics denote less error.

The energy norm and RMSE metrics served complimentary purposes. First note that the energy norm is a volume integration, whereas RMSE is a layer integration. Thus the former better represents the entire model simulation and was less sensitive to large errors in single layers. Second, the energy norm involves four variables (surface pressure, temperature, zonal wind, and meridional wind) and not just one, making its results more robust. Buizza et al. (2005) provided a compelling argument for the use of the energy norm as the primary validation metric and its current usage at ECMWF for model validation. Given all the above, the energy norm was the primary vehicle to evaluate model version simulation quality, and RMSE helped to identify error sources at levels throughout the troposphere.

5.3. Results

5.3.1 Simulated SST comparison

Figure 48 displays SSTs from case 7 for W321, WPol, and COAWST and the differences that arose by the end of each model run. This particular case was chosen because the associated SST gradients were the most distinct of the six cases. Domain 2 for each simulation was initialized with HYCOM within the ROMS domain (Fig. 47b) and RTG-SST elsewhere. By the model termination time (180 model hours later), SSTs are shown to be static for both W321 and WPol (Figs. 48d-e), yet they have evolved in the COAWST simulation (Fig. 48f). For both W321 and WPol, SSTs appear unchanged because the WRF SST update option was inactive for both simulations. Therefore SSTs were adjusted back to their initialization values each time the boundary conditions were updated and resulted in a 0 K SST difference. However, because the Pollard model altered the height of the oceanic mixed layer with time and ocean-based heat fluxes did change throughout the simulation unlike in W321. For COAWST model simulations SST variations at 180 hours ranged from -9.3K to 7.7K as compared to W321 and WPol simulations. Closer inspection of Figs 48f and 48i suggest these notable SST variations are attributed to a combination of wind-driven upwelling, shifts in the Gulf Stream itself (east of Virginia), and precipitation-based cooling of the immediate surface layer.

After running for 180 hours, SST and ocean-based heat flux changes had a notable impact upon simulated atmospheric fields. Altering just the mixed layer depth (i.e., WPol) produced point to point differences up to 3.8 hPa in SLP (case 8), 13.5 m s^{-1} (case 8) in 10 m winds, and 42 m in 500-hPa geopotential height as compared to W321. In comparison COAWST simulations produced point to point variations up to 6.0 hPa in

SLP (case 5), 14.5 m s^{-1} in 10 m winds (case 5), and 45 m in 500-hPa geopotential height. On average, meteorological field variability associated with the COAWST model was larger than that exhibited by WPol simulations. Comparing the average magnitude of these changes (model - W321) WPol variations were 89%, 89%, and 60% the size of COAWST variations in SLP, 10 m winds and 500-hPa geopotential height, respectively. Therefore despite being forced back to initialization SST values, meteorological field variability in response to ocean coupling in WPol is largely comparable to that seen with COAWST. Additionally, simulated field changes were not directly proportional to nor'easter cyclone intensity. Specifically, case 7, had the second strongest nor'easter (NESIS 4.38; minimum SLP 974 hPa), but it produced only the fourth largest SLP and 10 m wind variations. Finally, despite altering simulated meteorological fields to similar degrees, WPol and COAWST model simulation did not always agree. Taking the difference between simulations (COAWST – Wpol) revealed COAWST SLP and 10 m winds varied upwards of 7.18 hPa (case 5) and 12.4 m s^{-1} (case 7) from WPol, respectively.

5.3.2. WRF-ocean coupling analysis

Figure 49 depicts SLP-based storm tracks from all WRF simulations. Relative to GMA Fig. 49 displays storm track errors ranging from relatively minor (case 5) to extreme (case 4). To quantify track variability and ascertain potential biases, Fig. 50 displays GMA-relative track errors every six hours (small symbols) and their mean (larger symbols) from each WRF run. Shown errors are not latitude/longitude relative, but instead relative to the GMA cyclone propagation direction every six hours which

defines the positive y-direction. Using this framework, WRF track bias was found to range from 50 km (case 5) to over 2,400 km (case 4) and was typically within 150-250 km. Overall, 12 out of 18 (66.7%) WRF simulations exhibited a leftward track bias averaging 36 km and 10 out of 18 (55.5%) simulations lagged GMA averaging 61 km. Despite significant deviations from GMA, inter-WRF track differences averaged a mere 48 km, suggesting that similar to both Ren et al. (2004) and Yao et al. (2008) nor'easter storm tracks were not significantly altered by ocean-atmosphere coupling. To further quantify these biases Table 16 indicates the total number of six-hour periods where each individual model simulation exhibited a particular track bias relative to GMA. As seen in Table 16, all simulations types lagged GMA in at least 58.6% of all six hour periods and were nearly equally probable to exhibit either a leftward or rightward track bias.

Further analysis was completed to investigate the largest track errors which occurred in cases 4 and 8. Figure 51 displays various dynamical fields from case 4 on 12 UTC 28 February 2009 and 12 UTC 2 March 2009. These two times were selected to emphasize storm initialization and its later maturation. As seen in Figs. 51a-c and Figs. 51g-i, only minor differences exist between GMA and WRF, but 48 hours later (Figs. 51d-f and Figs. 51j-l) each WRF simulation varied little, but each differed impressively from GMA. Further investigation on 2 March found each WRF simulation maintained a 500-hPa cut-off height trough over Florida whereas GMA exhibits only a negatively tilted height trough. Due to this cut-off, simulated cyclone propagation speed fell to under 5 m s^{-1} in each WRF simulation. Eventually this slower propagation speed caused WRF simulations to have a vertically stacked cyclone, a weakened equivalent potential temperature ridge east of Florida, and weaker 500-hPa positive vorticity advection.

Combined these three factor helped to maintain the cut-off low three days longer than in GMA and produced WRF track errors in excess of 2,300 km and a 72 hour time lag.

Although not as impressive as case 4, case 8 storm track error was still notable (up to 1,100 km; W321). Figure 52 displays several key dynamical fields for case 8 on 00 UTC 14 March 2010. This time was selected because the differences between WRF and GMA were most apparent. Similar to case 4, a 500 hPa cut-off height trough was simulated by WRF (Figs. 52d-f), but unlike case 4 it also existed in GMA. Track error in this case however, originated with 500-hPa relative vorticity (Figs. 52d-f) and both the positioning of the 300-hPa jet (Figs. 52a-c), 850-hPa equivalent potential temperature ridge (Figs. 52g-i), and most importantly the positioning of the surface high (Figs 52a-c). All model runs simulated a nearly meridional 300-hPa jet while GMA showed this same jet to be nearly zonal. As a result, the surface cyclone over Michigan remained in the left-entrance jet region and under an area of overly strong positive vorticity advection which promoted its over-intensification (>10 hPa) and exaggerated 500-hPa height falls. Over-intensification of the Michigan cyclone modified geostrophic wind flow throughout the troposphere and positioned the equivalent potential temperature ridge too far eastward. Due to all the above factors, each model simulation was too aggressive with cyclone over the Eastern US (Figs. 52d-f) which preceded the nor'easter itself. As a result, each simulation failed to establish the strong (>1028 hPa) surface high over Nova Scotia. Because this surface high was not present. As seen in Figs. 52a-c, this error allowed each simulation to move the pre-nor'easter cyclone too far northward which then led to later errors with nor'easter cyclogenesis.

To assess simulated cyclone intensity accuracy, minimum SLP was compared between each WRF run and GMA. Simulated SLP was deemed “accurate” if it was within 5 hPa of GMA. Overall, model simulations over-intensified a nor’easter upwards of 11.9 hPa (case 1; WPol) and under-intensified by as much as 13.7 hPa (case 6; W321), which equates to 1.71 and -1.52 times the standard deviations of 7.82 hPa. All WRF simulations accurately predicted SLP in three cases (4, 7, and 8), over-intensified case 1, and under-intensified case 6. For case 5, WPol was the only simulation to over-intensify the nor’easter, but only did so by 0.23 hPa and had an SLP lower by up to 1.9 hPa (compared to W321) versus the other two simulations. This slight variability between simulations was associated with variations in atmospheric dynamics in response to ocean heat flux changes.

Investigating storm intensity variations involved an analysis of key meteorological fields (e.g., temperature, relative vorticity, latent heating, upper-level jet divergence). Figure 53 displays an analysis of 850-hPa equivalent potential temperature (Fig. 53a-b), 500-hPa geopotential height and 500-hPa positive relative vorticity (Fig. 53c-d), and 500-hPa latent heating (Fig. 53e-f) on 12 UTC 15 October 2009 for WPol and COAWST. This time was selected because all WRF simulations and GMA were within 50 km relative to each other, inter-WRF cyclone intensity varied less than 1 hPa, and synoptic-scale fields were similar. During the next six hours, minimum SLP fell 4.0 hPa and 4.5 hPa in WPol and COAWST, respectively. For both cases, positive vorticity advection near the cyclone center was small, but baroclinicity and latent heat release along the path led to pressure falls. Local changes in other parameters (not shown) were comparatively small. Figures 53a-b exhibit little to no difference in the simulated

equivalent potential temperature, but Figs. 53e-f do indicate slightly stronger mid-tropospheric latent heating just off the North Carolina coast in COAWST, which promoted additional upward vertical motion and pressure falls. For all other cases (not shown), latent heat release was strongly associated with SLP falls, but unlike in Fig. 53 it acted in concert with other favorable factors (e.g., positive vorticity advection) to produce the observed SLP variability.

Local-storm environment simulation accuracy was evaluated via energy norm differences ($W321 - \text{model}$) for both GMA- and WRF-centered frameworks (Fig. 54). The energy norm in each framework was evaluated within a 600-km-wide box centered upon on the GMA cyclone and each WRF cyclone, respectively. In Fig. 54, thin lines represent GMA-centered and thick lines represent WRF-centered energy norm differences and positive values denote improvement versus W321. Table 17 indicates the number of six-hour periods where WPol and COAWST simulations had a lower energy norm than W321. Overall, using ocean-atmosphere coupling improved up to 27/46 (58.7%; WPol) GMA-centered periods and 26/46 (56.5%; COAWST) WRF-centered periods. Individually, WRF-centered energy norm differences were positive in 50% or more periods in 1 and 3 out of 6 total cases for WPol and COAWST, respectively. In two cases (WPol, case 5; COAWST case 8) exactly 50% of periods were improved. Comparing COAWST to WPol (not shown) indicated COAWST improved 24/46 (52.2%) periods as compared to WPol. Additionally, the largest energy norm differences were often associated with cyclone propagation along the edge of the Gulf Stream otherwise energy norm differences in either coupled simulation relative to W321 were small.

To identify potential energy norm sources, WRF- and GMA-centered energy norms were time-averaged and horizontally-integrated at each vertical level (Fig. 55). In Fig. 55, the following three energy norm spikes were common to a majority of cases: ~850 hPa, ~650 hPa, and ~200 hPa. Further analysis (not shown) attributed these spikes errors in simulated planetary boundary layer height, the intensity and location of the nor'easter-driven mid-tropospheric jet (largest contributor), and tropopause height, respectively.

Energy norm differences integrated over the entirety of domains 1, 2 and 3 were used to assess the overall impact of ocean-atmosphere coupling on WRF simulations. To illustrate how energy norm differences varied between model domains and cases, Fig. 56 depicts these differences for cases 4 (Figs. 56a, 56c, 56e) and 5 (Figs. 56b, 56d, 56f). As seen in Fig. 56, energy norm differences from COAWST and WPol were often of the same sign and of similar magnitude with neither consistently exhibiting lower error than the other. Evaluation of time-averaged energy norms at each vertical level (not shown) was commensurate with the local-storm analysis, except that it exhibited a fourth energy norm spike around 300 hPa associated with the upper tropospheric jet stream. Energy norm contributions from 300 hPa were largest in domain 1 and smallest in domain 3, whereas contributions from 650 hPa followed the opposite trend. These two trends are associated with the spatial coverage, intensity, and location of the mid- and upper-tropospheric jets and their relationship to each domain. For example, the upper-level jet has large spatial coverage (approximately 1,000-3,000 km long and 500-1000 km wide), strong intensity (up to 150 knots), and is located in more poleward locations. As seen in Fig. 47a, a sizable region of domain 1 is located northward of 45°N, whereas domains 2

and 3 have only a smaller region north of this latitude. Therefore, the upper-level jet less frequently impacted domains 2 and 3 than domain 1, and thus it contributed less to the energy norms of these two domains. In contrast, although the mid-tropospheric jet is of lower magnitude (<80 knots) and smaller spatial extent (500-1000 km long and 300-500 km wide), it was often associated with all nor'easter cases. Because of these attributes and its association with nor'easters, the mid-tropospheric jet had modest error contributions on domain 1, yet larger contributions for domains 2 and 3.

As compared to W321, coupled ocean-atmosphere nor'easter simulations exhibited lower energy norms for at most 101 (54.3%; COAWST), 101 (54.3%; COAWST), and 99 (53.2%) out of 186 six-hour periods for domains 1, 2 and 3, respectively (Table 17). For domains 1, 2, and 3 WPol simulation had a lower energy norm than W321 in 2, 1, and 3 out of 6 cases, respectively and in as many as 91 (48.9%) out of 186 periods. Simulations from COAWST had a lower energy norm than W321 in 4, 4, and 5 cases, respectively and in as many as 101 (54.3%) out of 186 periods. Thus COAWST model simulations improved 20% more periods than WPol as compared to W321. Direct comparison between COAWST and WPol revealed the former had a lower energy norm in up to 109 (58.6%; domain 1) of 186 periods and for 3, 4, and 3 out of 6 cases, respectively. Despite these results, the average energy norm differences from W321 was on the order of 10^8 J m s^{-2} , which although large is still 2-3 orders of magnitude less than the energy norm itself. Given earlier results (Figs. 51-53) energy norm differences of this scale likely do not indicate a significant impact to overall simulation outcome. To illustrate the model impact of ocean coupling, Fig. 57 shows SSTs and SLP for GMA and the three model simulations on 00 UTC 17 October 2009.

As seen in Fig. 57, the overall SLP field between model runs varied only slightly despite the northward shift in SST south of Nova Scotia. This northward shift in SSTs however did cause the COAWST nor'easter to exhibit a slightly more poleward track as compared to W321 and WPol seen earlier in Fig. 49 (case 5).

For comparison to the energy norm results and to determine which dynamical fields were most sensitive to WRF-ROMS coupling, a series of RMSE calculations were completed and summarized in Table 17. The five RMSE error variables shown in Table 17 were selected either due to their strong emphasis in Kocin and Uccellini (2004) or as an indicator of the dynamic tropopause (i.e., 2-PVU potential temperature). Values shown in Table 17 reflect the number of six-hour periods where a particular model has a lower RMSE value than W321. With the exception of WPol 500-hPa geopotential height, Table 17 indicated the number of improved RMSE time periods were largely consistent with the energy norm analysis. How these differences varied with time however, was not always consistent either with the energy norm or between the various RMSE metrics. Figure 58 displays differences (W321 - model) in energy norm (Fig. 58a) and the five RMSEs (Figs. 58b-e) for case 1. Case 1 was selected for its representativeness of the entire sample. Consistent with Yao et al. (2008), atmospheric adjustments to the coupled ocean-atmosphere model extended at least as high as 500 hPa as evidenced by SLP, 850-hPa temperatures and 500-hPa geopotential height RMSEs following similar patterns and changing sign in sync. Other RMSEs did not follow these patterns suggesting little to no impact from model coupling at these levels.

5.4. Conclusions

The impact of WRF-ocean model coupling on nor'easter simulation accuracy was evaluated. All model simulations were configured identically, run for 180 model hours, initialized 72 hours prior to first precipitation impacts in the highly populated Mid-Atlantic US, and evaluated using GMA. They differed only in how each handled the ocean. For each of the six nor'easter cases an uncoupled W321, a coupled WRF-Pollard 1D ocean mixed layer, and a coupled COAWST (WRF-ROMS coupling) simulation were completed. Over the 180-hour runtime, adjustments to the total heat flux altered SLP, 10 m wind, and 500-hPa geopotential height by up to 6.0 hPa, 14.5 m s⁻¹, and 45 m, respectively. Given that the magnitude of WPol associated atmospheric field changes were generally 60% or more that associated with COAWST suggests the importance of heat flux variations originating from the oceanic mixed layer. Other factors such as horizontal advection, shifts in the Gulf Stream itself, and precipitation-based cooling of the immediate surface layer may have a smaller, but still notable impact. Coupling WRF to either ROMS or Pollard models produced little to no changes in storm track consistent with Ren et al. (2004), Yao et al. (2008) and Warner et al. (2010). A majority of WRF simulations (12 out of 18) favored a leftward bias and (10 out of 12) lagged GMA. While no WRF simulation exhibited a defined intensity bias, COAWST simulated minimum SLP values slightly better than both other models. Inter-WRF comparison of dynamical fields attributed minimum SLP differences to variations in mid-tropospheric latent heat release.

WRF- and GMA-centered energy norm differences were improved over W321 in up to 27 (58.7%) and 26 (56.5%) out of 46 total six-hour periods, respectively and in up

to 5 out of 6 cases. Energy norm differences were often close to zero suggesting each simulation exhibited only minor simulation changes and accuracy improvement. Time- and horizontally-averaged energy norm profiles detected four distinct energy norm sources associated with the planetary boundary layer height and the low-level jet, mid-tropospheric jet intensity and positioning, upper-level jet stream intensity and positioning, and tropopause height.

Energy norm and RMSE calculations for domains 1, 2, and 3 permitted investigation of how ocean-atmosphere coupling impacted the simulation holistically. Energy norm results for WPol and COAWST improved WRF simulations in up to 91 (48.9%) and 101 (54.3%) out of 186 periods, respectively. Over all 3 model domain, only COAWST outperformed W321 in more than 50% of all time periods. Most energy norm differences however, were 2-3 orders of magnitude less than the energy norm which suggests only modest simulation modifications. Analysis of RMSEs found a similar number of time periods where ocean coupling exhibited improvement compared to W321, but results above 500 hPa was likely unassociated with ocean-atmosphere coupling consistent with Yao et al. (2008).

Overall, this study demonstrated that while coupled ocean-atmosphere models may indeed produce more realistic ocean and atmosphere simulations, they did not lead to significant improvements in overall simulation accuracy. While overall simulation accuracy may not have improved greatly, atmosphere ocean-coupling can still significantly alter SSTs, heat fluxes, SLP and surface winds among others. This suggests that other more localized quantities such as precipitation, convection, and low-level cloud cover may also show significant differences between coupled and uncoupled simulations.

Finally, results from Warner et al. (2010) found that storm track was also not notably altered when coupling WRF to ROMS, but when coupling both of these models to SWAN storm track was seen to deviate. Thus the impact of waves and their breaking upon nor'easters may be more noticeable and merits future study.

5.5. Future work

This work primarily investigated how model accuracy and various dynamic fields were altered by ocean-atmosphere coupling, but much more expansion is possible. As shown in Fig. 46, this study only coupled WRF to ROMS via COAWST, but Warner et al. (2010) has demonstrated that surface stress values are notably altered with the inclusion of SWAN for tropical cyclones. Would this also be true for nor'easters? Model verification in this study strictly involved GMA, but future work could utilize satellite and ground-based data to evaluate simulations of SST, precipitation, or cloud cover. Next, other than for SSTs no analysis involved the ocean. Future work could verify model output against Coastal Ocean Dynamics Applications Radar (CODAR) data to evaluate simulated ocean currents. Finally, the WRF model domain used in this research had 15 km grid spacing. Would adjusting this value closer to the ROMS grid produce better results?

Table 15: Nor'easter case list. Column 3 shows the dates which each event impacted the northeastern US, while the last two columns denote the first and last times for each model run. Due to missing HYCOM data, cases 2 and 3 are not analyzed.

Case Number	NESIS	Event Dates	Model Run Start Date	Model Run End Date
1	N/A	22-24 Nov. 2006	11/19 12UTC	11/27 00UTC
2	2.54	15-17 Mar. 2007	3/12 18UTC	3/20 06UTC
3	N/A	15-17 Apr. 2007	4/12 06UTC	4/19 18UTC
4	1.65	1-2 Mar. 2009	2/26 12UTC	3/6 00UTC
5	N/A	15-16 Oct 2009	10/12 12UTC	10/20 00UTC
6	4.03	19-20 Dec. 2009	12/16 06UTC	12/23 18UTC
7	4.38	4-7 Feb. 2010	2/2 18UTC	2/10 06UTC
8	N/A	12-14 Mar. 2010	3/9 06UTC	3/16 18UTC

Table 16: GMA-relative storm track bias analysis. Values indicate the total the number six-hour time periods from all cases where each simulation exhibited the shown track bias relative to GMA at each time period.

	Left		Right		Key	
Lead	11	9	12	15		
	9		11		W321	Wpol
Lag	17	20	16	14	COAWST	
	21		16			

Table 17: Performance analysis relative to W321. Values denote the number and percentage of 6-hour periods where the energy norm or RMSE from the indicated simulations was lower than W321. The last section indicates the number of periods where WRF model run energy norms were lower than W321 within a 600-km box centered on GMA and each simulated storm, respectively. Bolded values indicate the best WRF simulation (as compared to W321) for each analysis component.

Domain 1 Performance Analysis vs W321 (186 Total Times)	Wpol	COAWST
Energy Norm	83 (44.6%)	101 (54.3%)
RMSE 2 PVU Pot T	75 (40.3%)	84 (45.2%)
RMSE 300-hPa Winds	92 (49.5%)	99 (53.2%)
RMSE 500-hPa Geo Hght	93 (50.0%)	90 (48.4%)
RMSE 850-hPa Temps	87 (46.8%)	99 (53.2%)
RMSE SLP	85 (45.7%)	94 (50.5%)
Domain 2 Performance Analysis vs W321 (186 Total Times)	Wpol	COAWST
Energy Norm	79 (42.5%)	101 (54.3%)
RMSE 2 PVU Pot T	87 (46.8%)	113 (60.8%)
RMSE 300-hPa Winds	88 (47.3%)	92 (49.5%)
RMSE 500-hPa Geo Hght	93 (50.0%)	86 (46.2%)
RMSE 850-hPa Temps	67 (36.0%)	87 (46.8%)
RMSE SLP	86 (46.2%)	96 (51.6%)
Domain 3 Performance Analysis vs W321 (186 Total Times)	Wpol	COAWST
Energy Norm	91 (48.9%)	99 (53.2%)
RMSE 2 PVU Pot T	93 (50.0%)	102 (54.8%)
RMSE 300-hPa Winds	86 (46.2%)	81 (43.5%)
RMSE 500-hPa Geo Hght	95 (51.1%)	90 (48.4%)
RMSE 850-hPa Temps	73 (39.2%)	96 (51.6%)
RMSE SLP	85 (45.7%)	100 (53.8%)
Domain 2 Storm Performance Analysis vs WRF 321 (46 Total Times)	Wpol	COAWST
Energy Norm GFS Rel	27 (58.7%)	25 (54.3%)
Energy Norm WRF rel	20 (43.5%)	26 (56.5%)

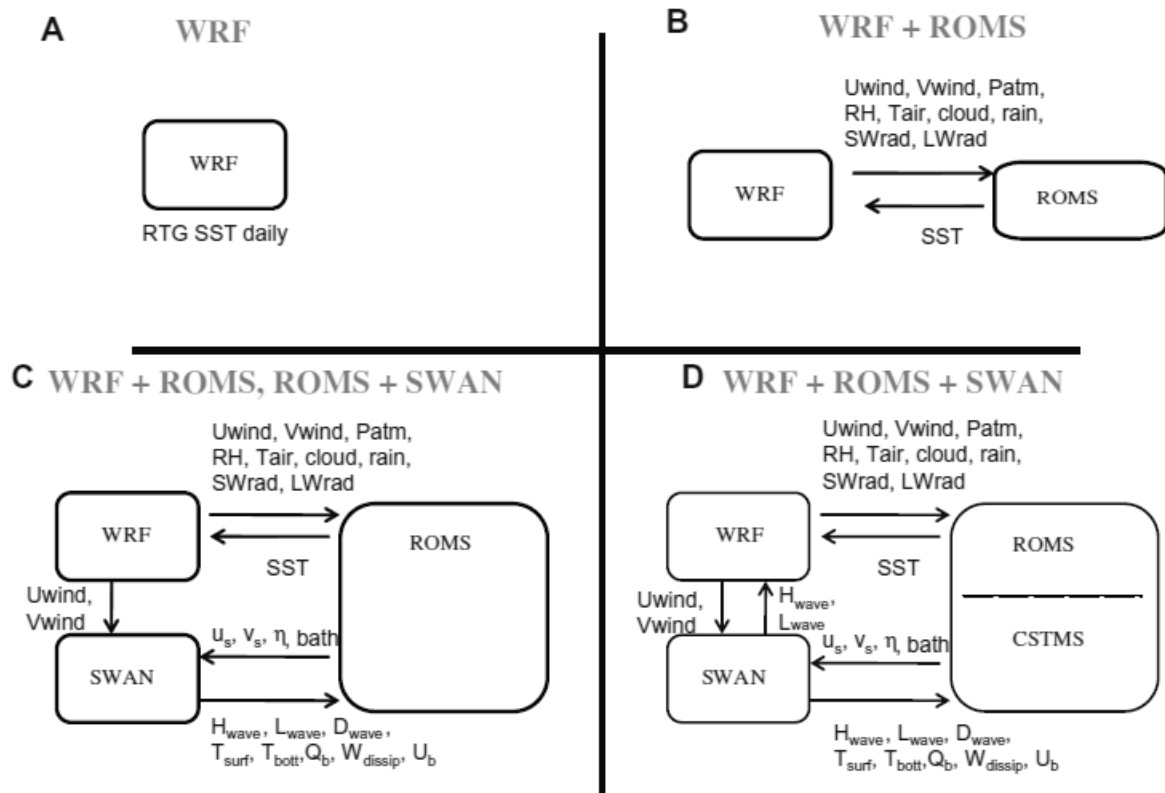


Fig. 46: All available COAWST configurations involving the WRF model and exchanged data fields. (a) Is WRF only, (b) WRF-ROMS coupling, (c) WRF-ROMS-SWAN coupling, (d) WRF-ROMS-SWAN-CSTMS coupling. Adapted from Fig. 5 of Warner et al. (2010).

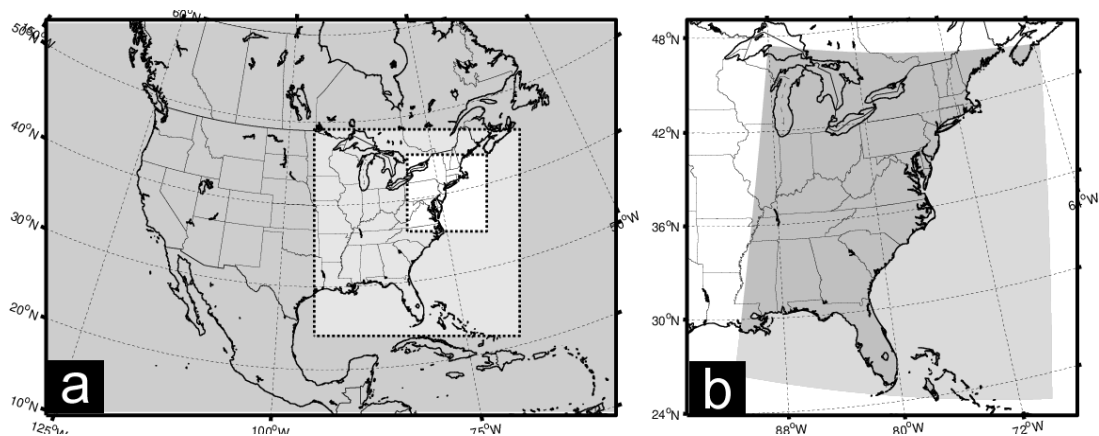


Fig. 47: (a) WRF model configuration where domains 1, 2, and 3 have 45, 15, and 5 km grid spacing, respectively. (b) ROMS model configuration superimposed on WRF model domain 2 with 8 km average grid spacing.

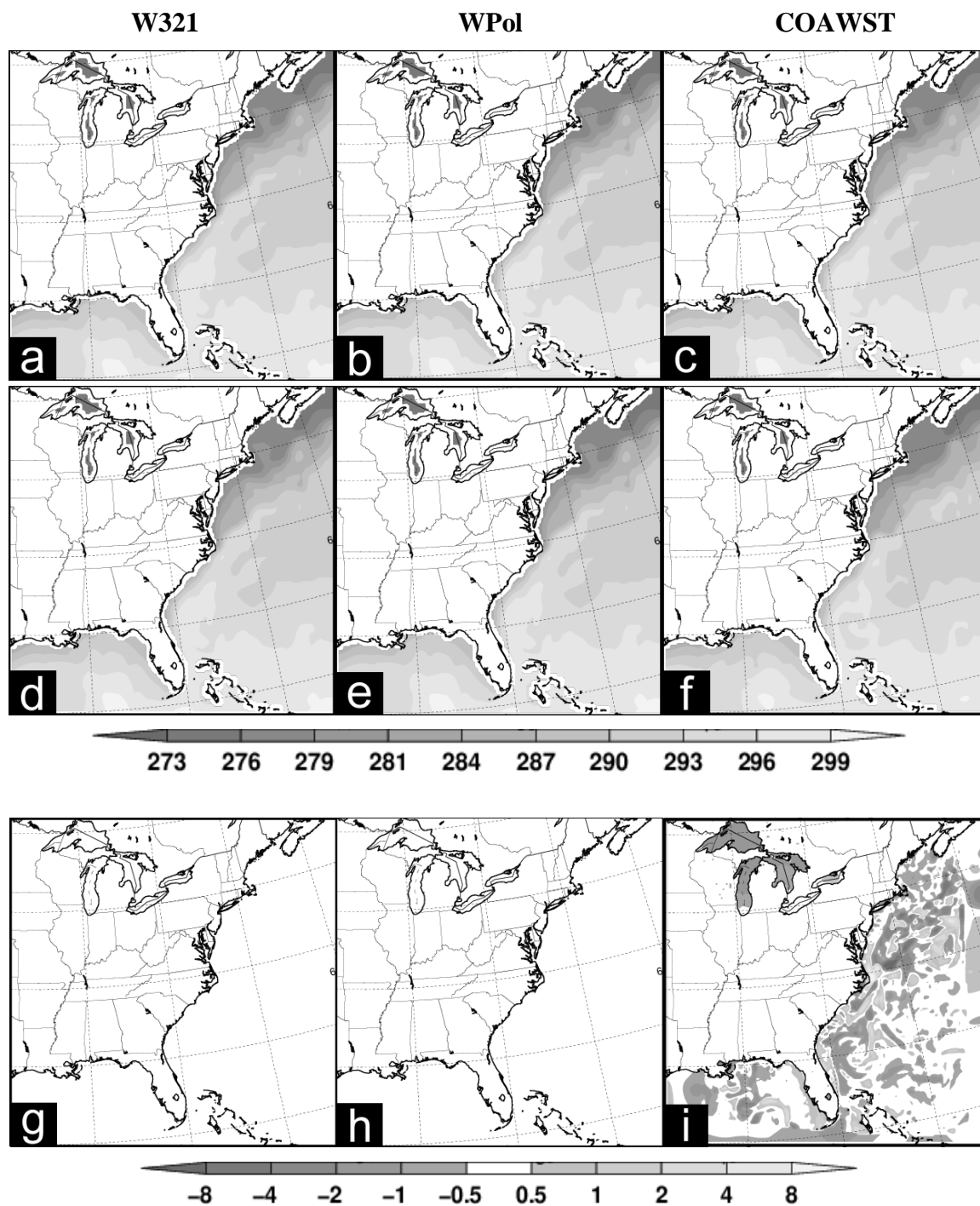


Fig. 48: Simulated sea surface temperatures (K) from W321, WPol, and COAWST at (a-c) 18UTC 2 February 2010, (d-f) 06UTC 10 February 2010 and (g-i) their difference.

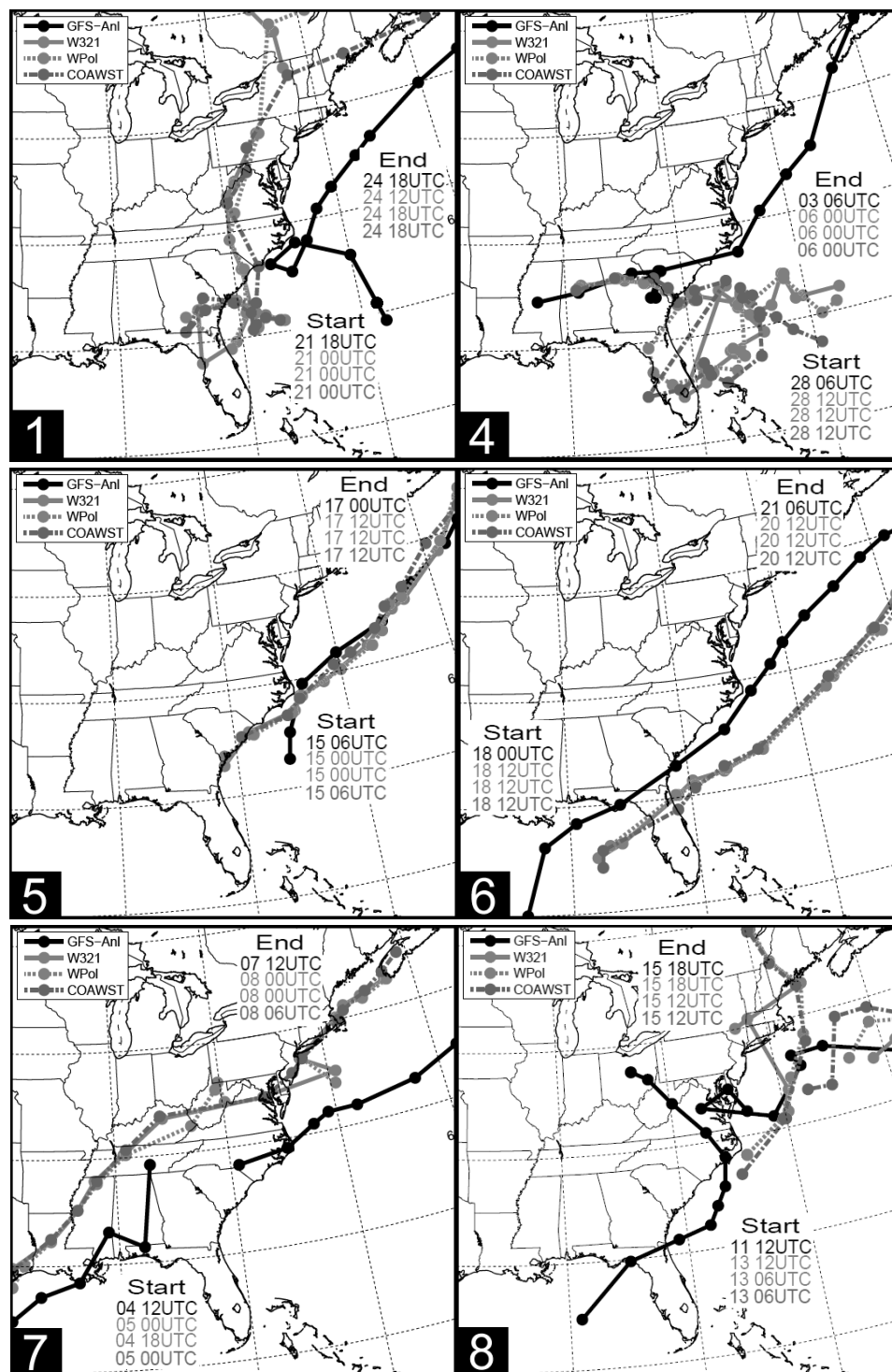


Fig. 49: Storm tracks from GMA and WRF runs. Line legend is shown on the upper-left of each plot. Shown symbols indicate simulated storm position every six hours. White numbers indicate case number.

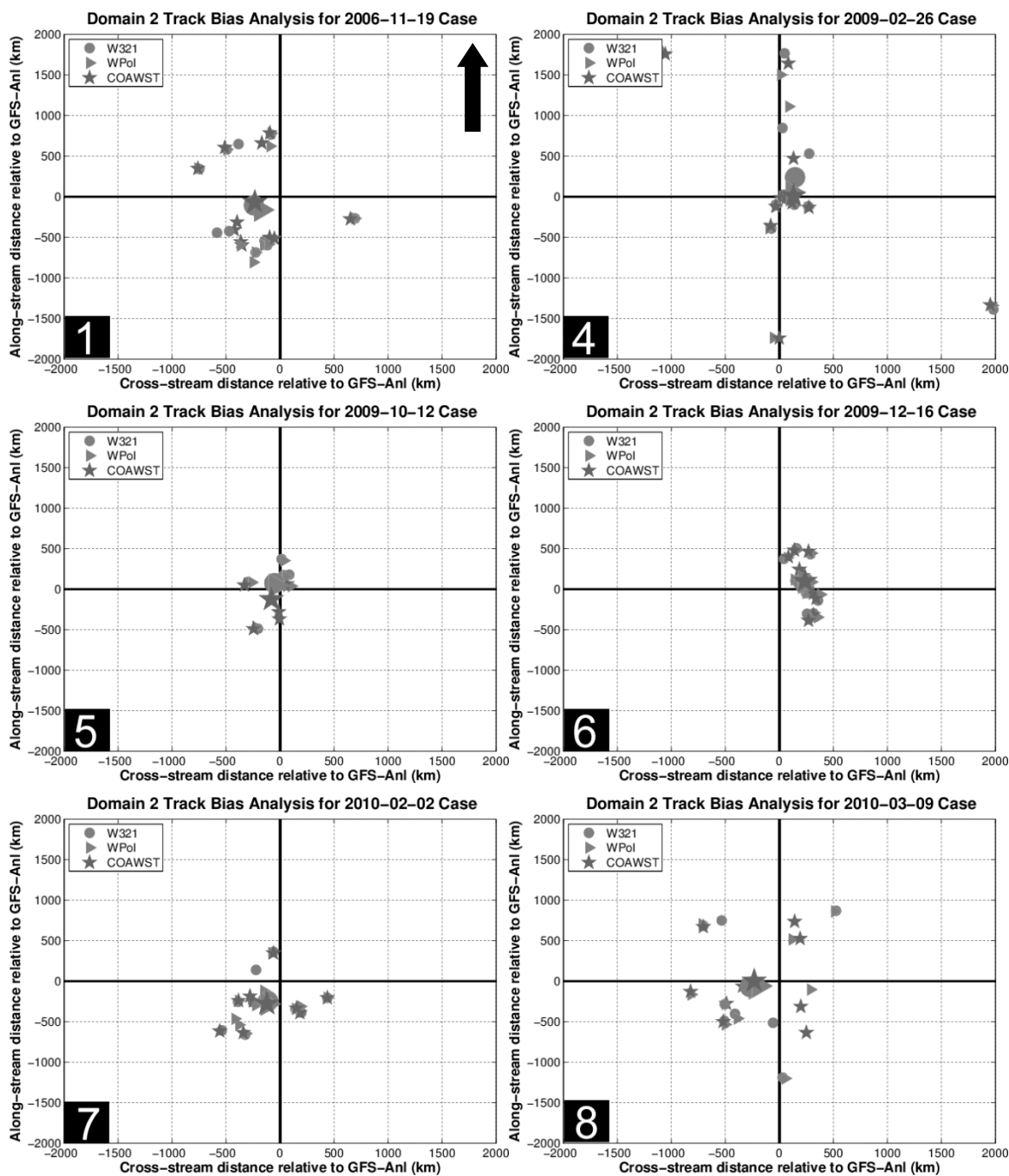


Fig. 50: WRF forecasted storm position bias as compared to GMA for all eight cases. The black arrow in panel 1 indicates the GMA storm motion direction for all panels. Shown symbols represent WRF position bias every six hours (smaller symbols) and their mean (large symbols). Case number is indicated with a white number.

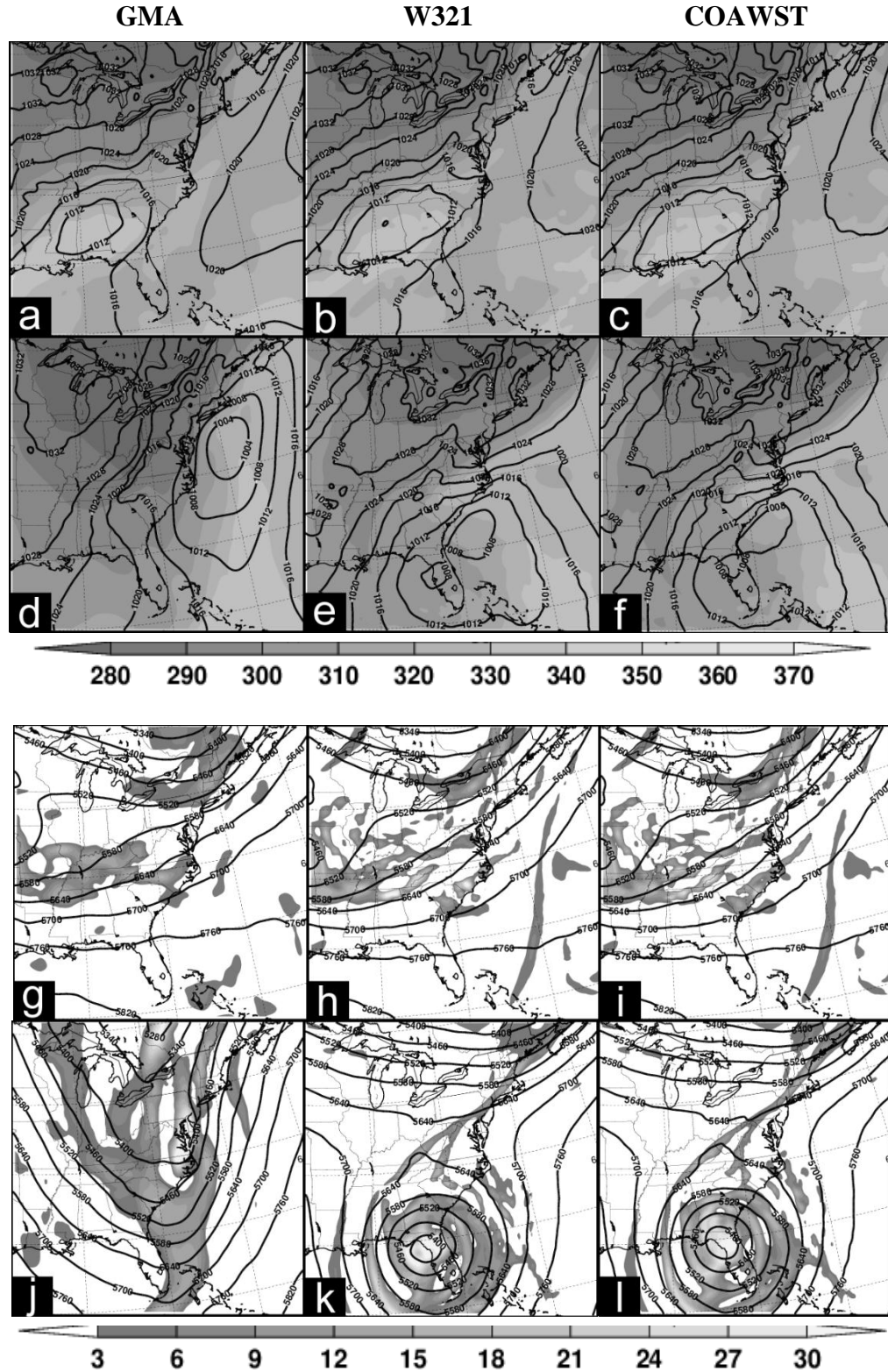


Fig. 51: Various plots from GMA, W321 and COAWST. (a-f) 850-hPa equivalent potential temperature (fills, K) and sea-level pressure (contours) and (g-l) 500-hPa positive relative vorticity (fills, 10^{-5} s^{-1}) and 500-hPa geopotential height (fills, m). (a-c, g-i) are on 28 February 2009 12UTC and (d-f, j-l) are on 12 UTC 2 March 2009.

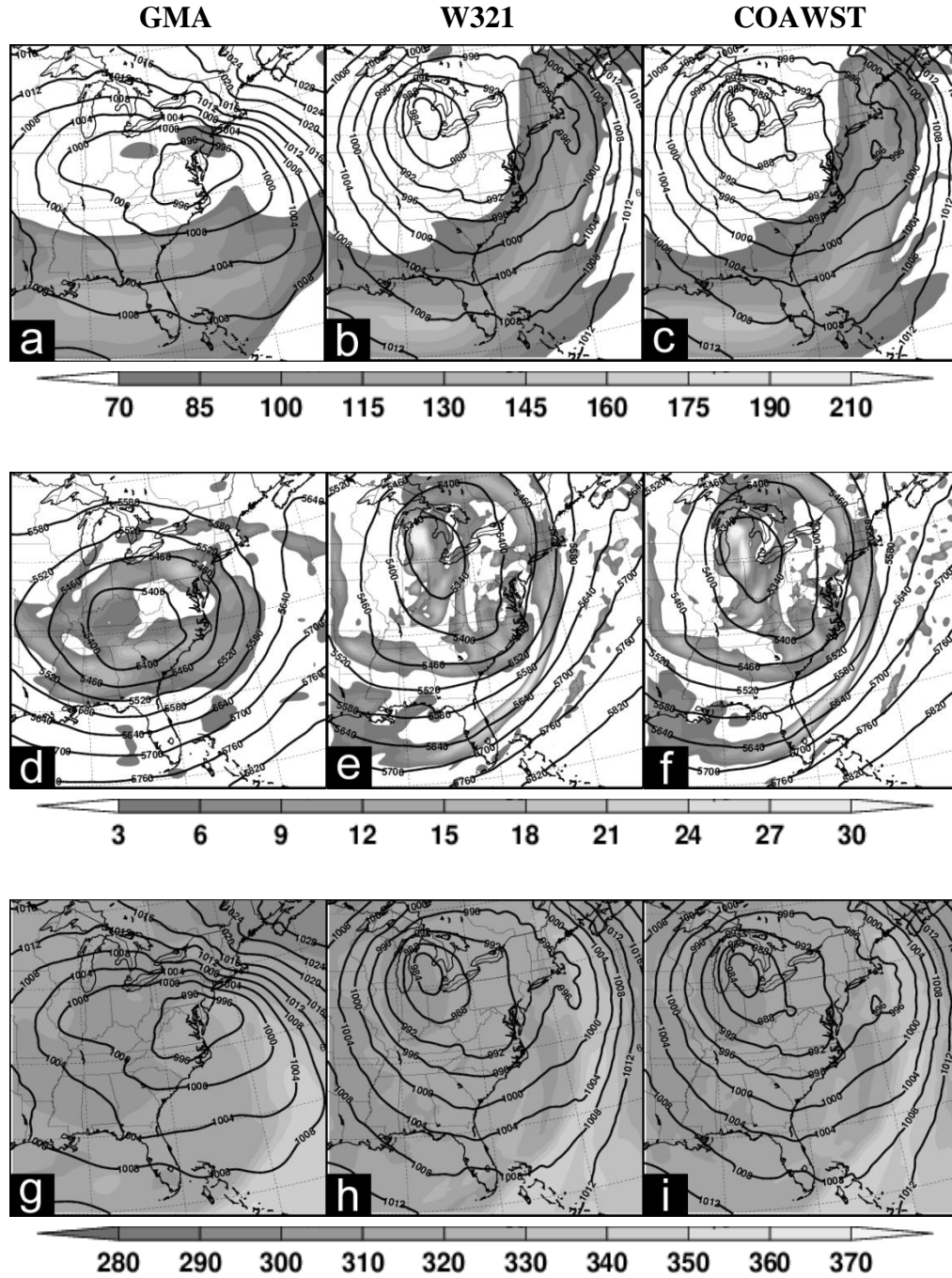


Fig. 52: Various plots from GMA, W321 and COAWST. (a-c) 300-hPa wind speed (knots, fills) and sea-level pressure (hPa, contours), (d-f) 500-hPa positive relative vorticity (fills, 10^{-5} s^{-1}) and 500-hPa geopotential height (fills, m), and (g-i) 850-hPa equivalent potential temperature (K, fills) and sea-level pressure (hPa, contours). All plots are for 00UTC 14 March 2010.

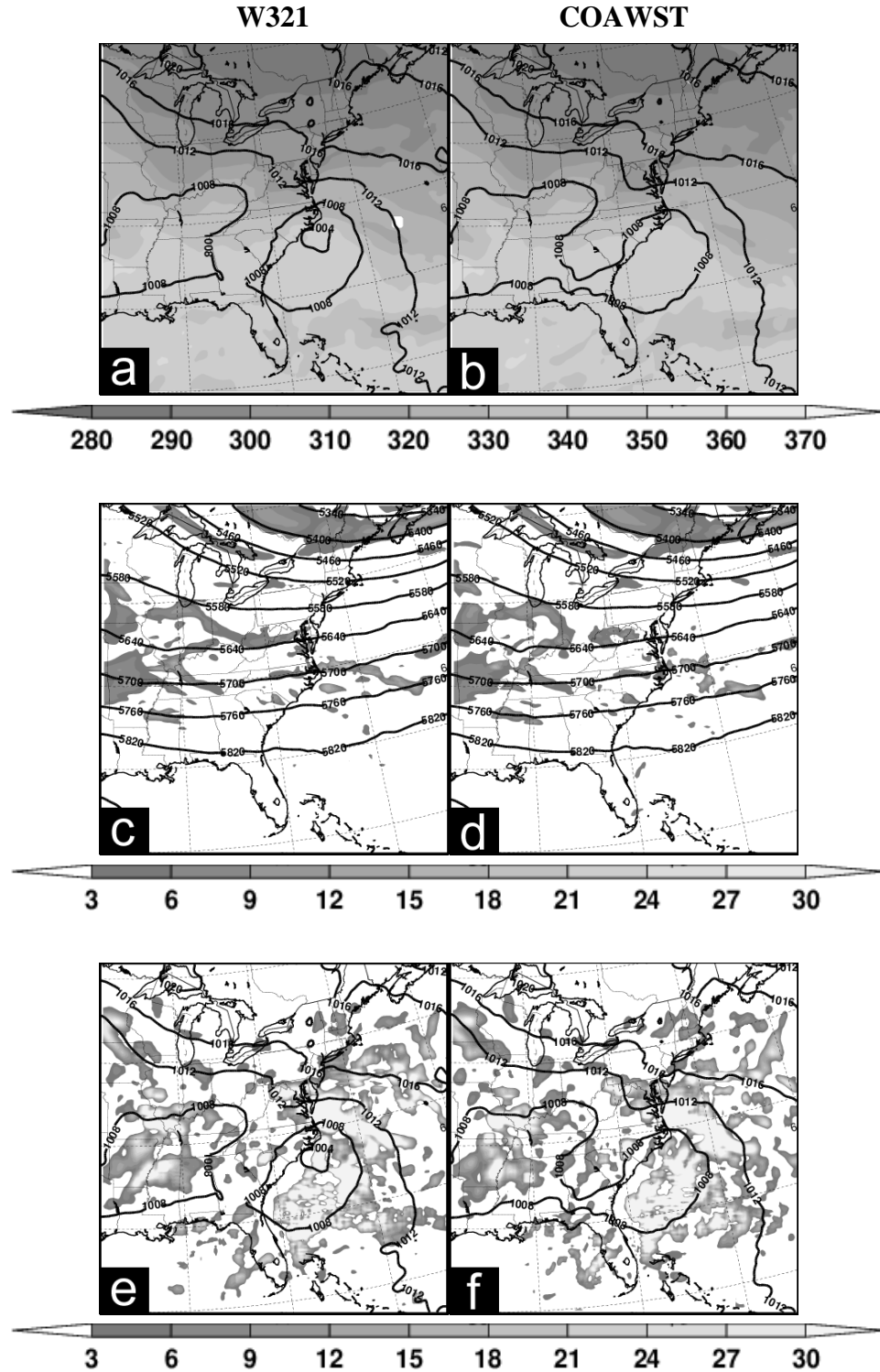


Fig. 53: (a-b) 850-hPa equivalent potential temperature (fills, K) and sea-level pressure (hPa, contours), (c-d) 500-hPa positive relative vorticity (fills, 10^{-5} s^{-1} and 500-hPa geopotential height (m, contours), and (e-f) 500-hPa latent heating (fills, K day^{-1}) and sea-level pressure (contours, hPa). All plots are from 15 October 2009 at 12UTC.

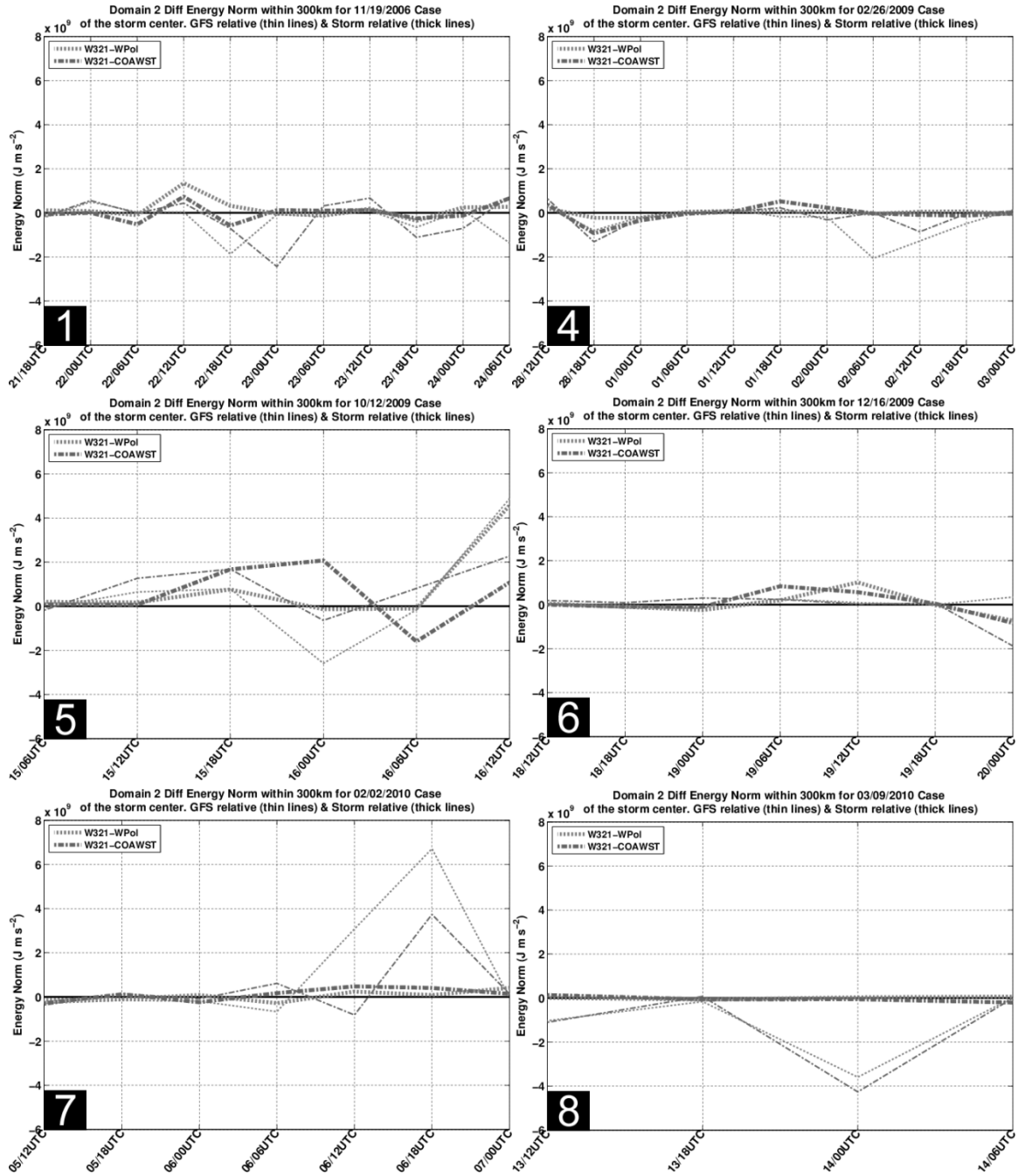


Fig. 54: Domain 2 energy norm differences for all cases within 300 km of the GMA storm center (thin lines) and each model simulated center (thick lines). Shown differences are relative to W321 with positive values denoting improvement.

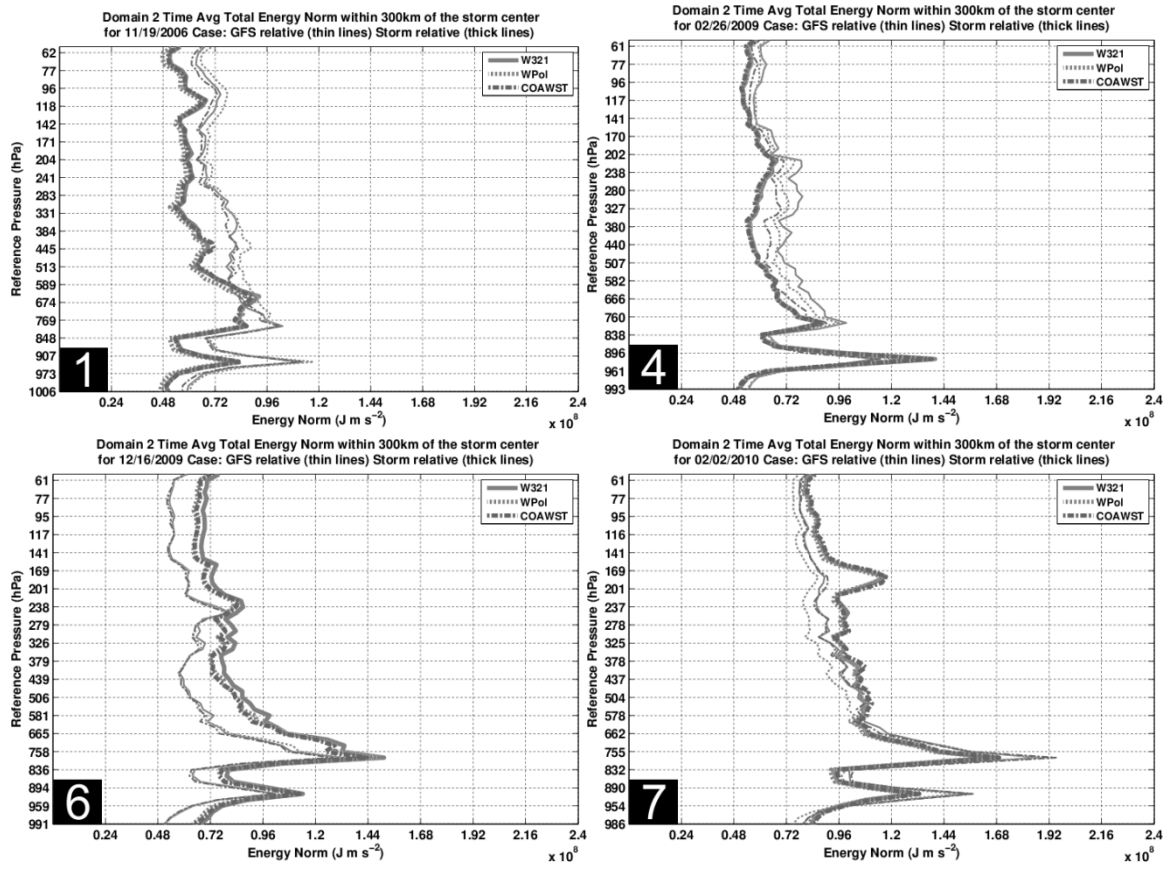


Fig. 55: Domain 2, time-averaged energy norm for cases 1, 4, 6 and 7 with 300 km of the GFS analysis storm center (thin lines) and each model simulated center (thick lines).

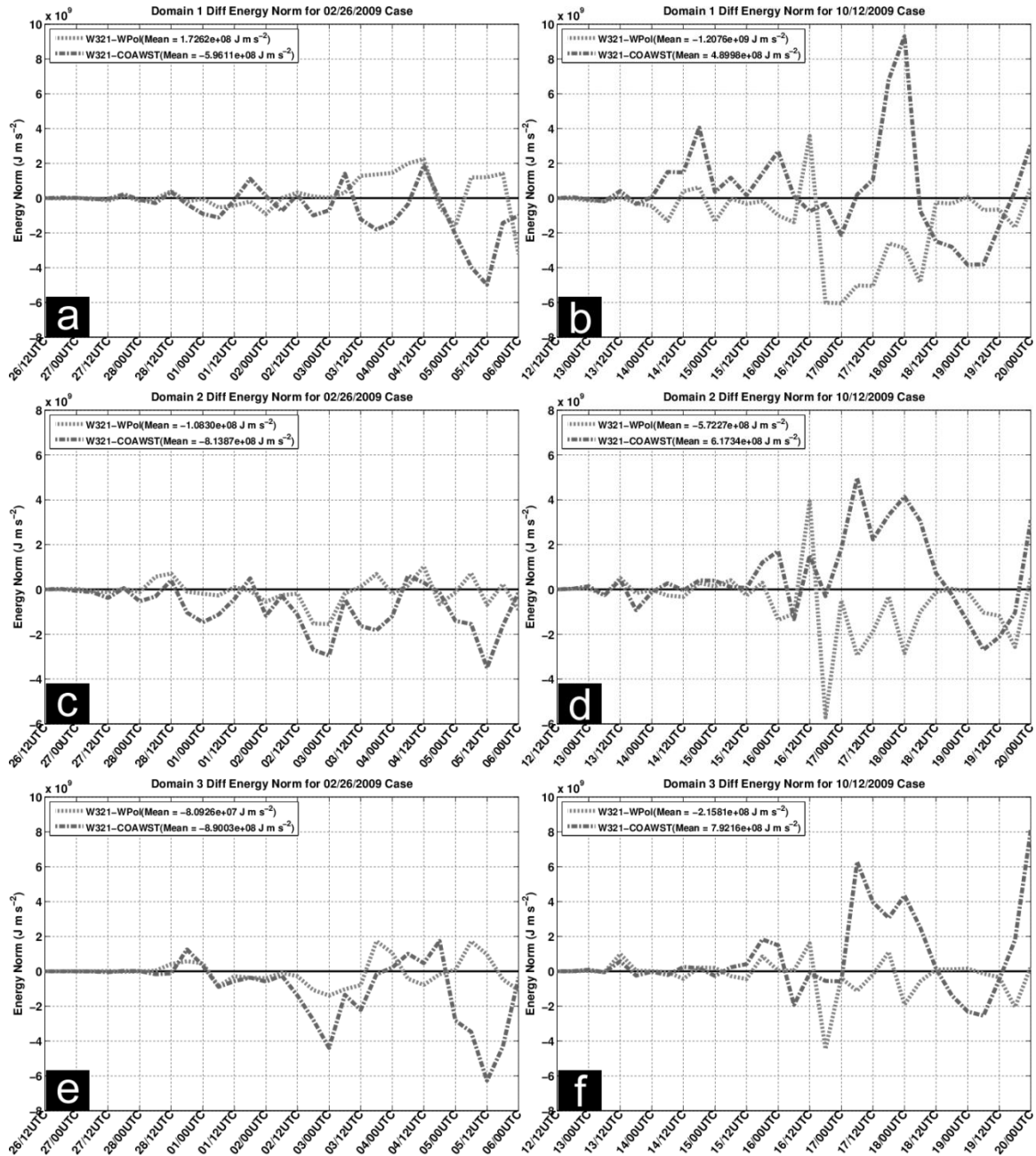


Fig. 56: Energy norm differences within the entirety of domains 1, 2, and 3 for (a, c, e) case 4 and (b, d, e) case 5. Shown differences are relative to W321. Positive values denote improvement over W321.

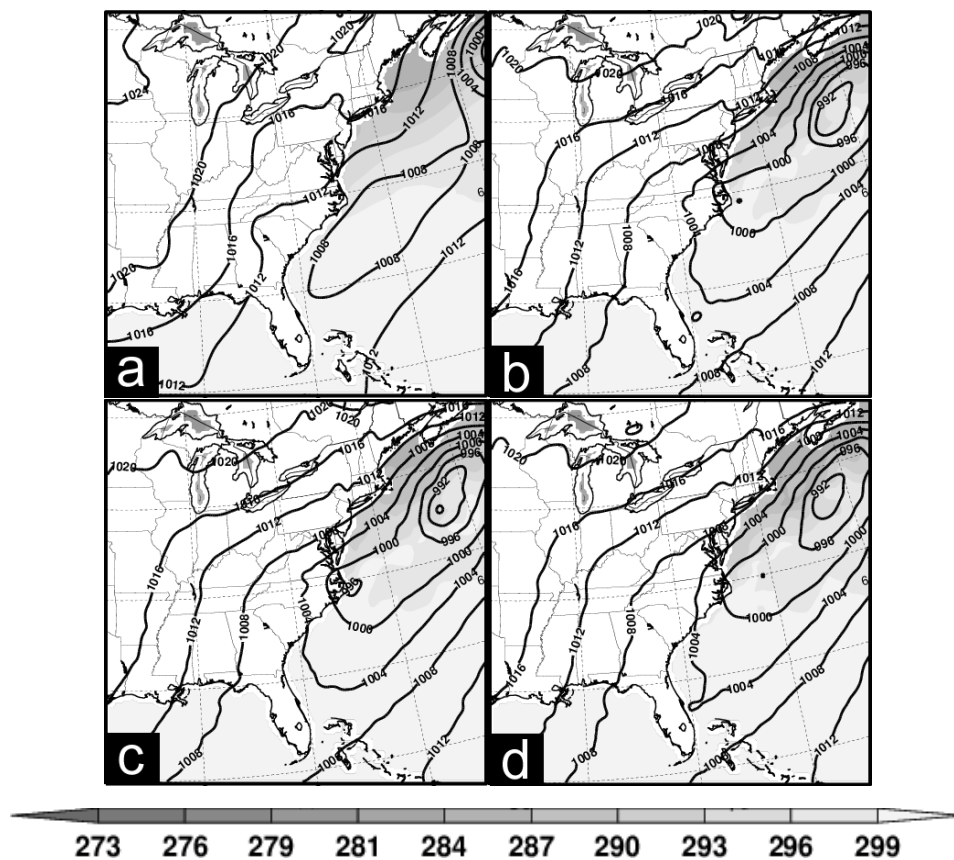


Fig. 57: SST (fills, K) and SLP (contours, hPa) on 00 UTC 17 October 2009 from (a) GMA, (b) W321, (c) WPol, and (d) COAWST.

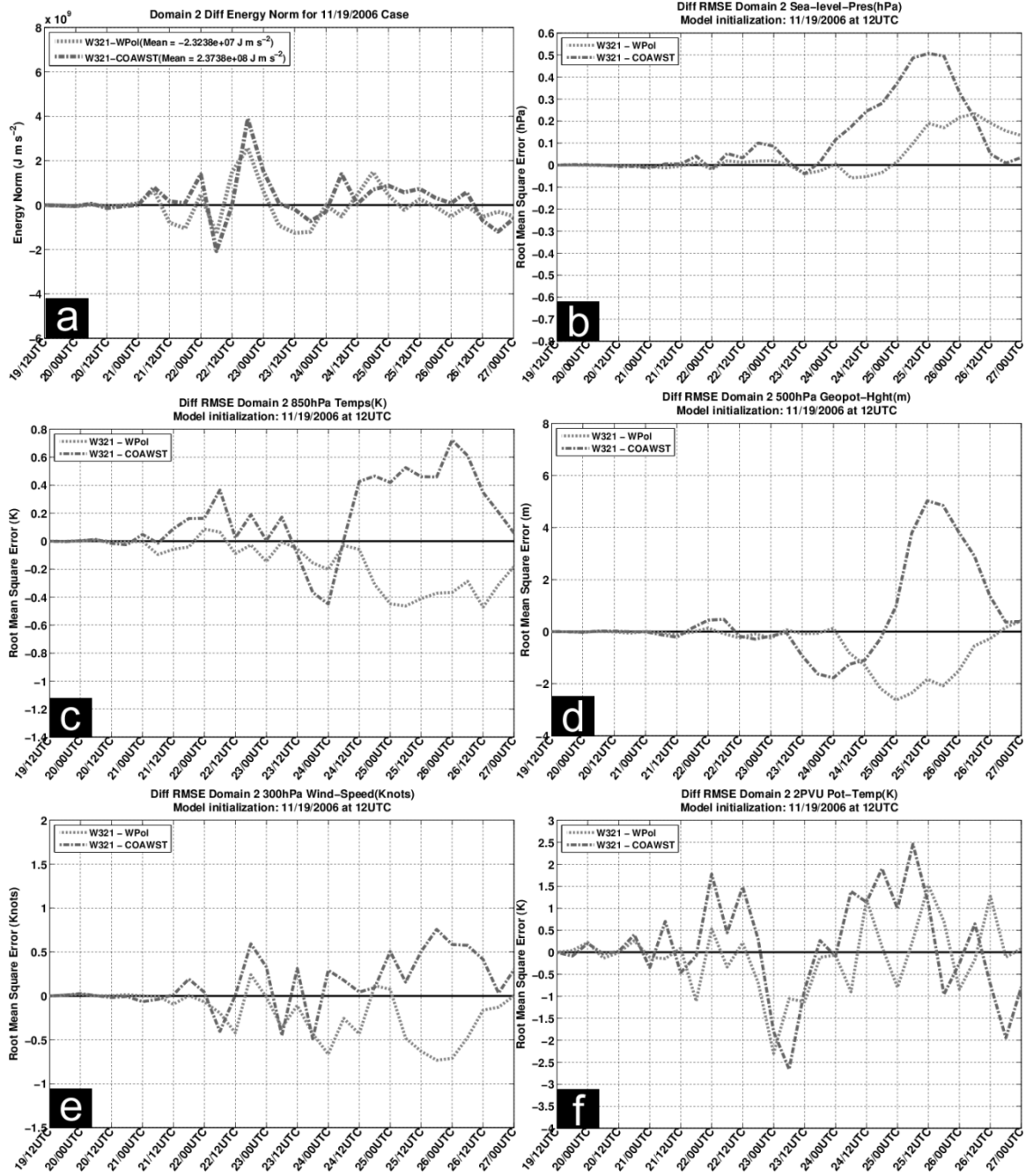


Fig. 58: Difference in domain 2 (a) energy norm, (b) sea-level pressure, (c) 850-hPa temperature, (d) 500-geopotential height, (e) 300-hPa winds, (f) 2-PVU potential temperature for case 1. Shown differences are relative to W321. Positive values denote improvement over W321.

Bibliography

- Anthes, R. A., Y. -H. Kuo, and J. R. Gyakum, 1983: Numerical simulations of a case of explosive marine cyclogenesis, *Mon. Wea. Rev.*, **111**, 1174-1188.
- Anthes, R.A. and Coauthors, 2008: The COSMIC/FORMOSAT-3 Mission: Early Results. *Bull. Amer. Meteor. Soc.*, **89**, 313–333.
- Anthes, R. A., 2011: Exploring Earth's atmosphere with radio occultation: contributions to weather, climate and space weather. *Atmos. Meas. Tech.*, **4**, 1077-1103, doi:10.5194/amt-4-1077-2011.
- Aragon-Angel, A., Y. -A. Liou, C. -C. Lee, B. W. Reinisch, M. Hernández-Pajares, M. Juan, and J. Sanz, 2011: Improvement of retrieved FORMOSAT-3/COSMIC electron densities validated by ionospheric sounder measurements at Jicamarca. *Radio Sci.*, **46**, RS5001, doi:10.1029/2010RS004578.
- Ashton, A.D., J.P. Donnelly, and R.L. Evans, 2008: 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.
- Barker, D. M., W. Huang, Y. R. Guo, and Q. N. Xiao, 2004: A three-dimensional (3DVAR) data assimilation system for use with MM5: Implementation and initial results. *Mon. Wea. Rev.*, **132**, 897–914.
- Bender, M. A., I. Ginis, and Y. Kurihara, 1993: Numerical simulations of tropical cyclone-ocean interaction with a high-resolution coupled model. *J. Geophys. Res.*, **38**, 23,245–23,263.
- Biondi, R., T. Neubert, S. Syndergaard, and J. K. Nielsen, 2011: Radio occultation bending angle anomalies during tropical cyclones, *Atmos. Meas. Tech.*, **4**, 1053-1060.
- Boniface, K., J. M. Aparicio, and E. Cardellach, 2011: Meteorological information in GPS-RO reflected signals. *Atmos. Meas. Tech.*, **4**, 1397-1407, doi:10.5194/amt-4-1397-2011.
- Booij, N., Ris, R. C., Holthuijsen, L. H., 1999: A third-generation wave model for coastal regions. Part I: Model description and validation. *J. Geophys. Research*, **104**, 7649–7666.
- Bougeault, P., and P. Lacarrère, 1989: Parameterization of orography-induced turbulence in a mesobeta-scale model. *Mon. Wea. Rev.*, **117**, 1872-1890.
- Brown, A. R., and Coauthors, 2002: Large-eddy simulation of the diurnal cycle of shallow cumulus convection over land. *Quart. J. Roy. Meteor. Soc.*, **128**, 1075–1093.
- Buizza, R., P. L. Houtekamer, G. Pellerin, Z. Toth, Y. Zhu, and M. Wei, 2005: A Comparison of the ECMWF, MSC, and NCEP Global Ensemble Prediction Systems. *Mon. Wea. Rev.*, **133**, 1076–1097.
- Cavallo, S. M., J. Dudhia, C. Snyder, 2010: A multi-layer upper boundary condition for longwave radiative flux to correct temperature biases in a mesoscale model. *Mon. Wea. Rev.*, **139**, 1952-1959.
- Chelton, D. B., and Coauthors, 2001: Observations of coupling between surface wind stress and sea surface temperature in the eastern tropical Pacific. *J. Climate*, **14**, 1479–1498.
- Chelton, D. B., 2005: The impact of SST specification on ECMWF surface wind stress fields in the eastern tropical Pacific. *J. Climate*, **18**, 530-550.

- Chen, F., and J. Dudhia, 2001: 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.
- Chen, F., Kusaka, H., Bornstein, R., and Coauthors, 2011: The integrated WRF/urban modelling system: development, evaluation, and applications to urban environmental problems. *Int. J. Climatol.*, **31**, 273–288.
- Cheng, C.-Z., Y.-H. Kuo, R. A. Anthes, and L. Wu, 2006: Satellite constellation monitors global and space weather, *Eos Trans. AGU*, **87**, doi:10.1029/2006EO170003.
- Chou, M.-D., and M. J. Suarez, 1999: A solar radiation parameterization for atmospheric studies. NASA Tech. Rep. NASA/TM-1999-10460, vol. 15, 38pp.
- Chou, M.-D., and M. J. Suarez, 2001: A thermal infrared radiation parameterization for atmospheric studies. NASA Tech. Rep. NASA/TM-1999-10466, vol. 19, 55pp.
- Cucurull, L., J. C. Derber, R. Treadon, and R.J. Purser, 2008: Preliminary impact studies using Global Positioning System radio occultation profiles at NCEP. *Mon. Wea. Rev.*, **136**, 1865–1877.
- Dudhia, J., 1989: Numerical study of convection observed during the winter monsoon experiment using a mesoscale two-dimensional model, *J. Atmos. Sci.*, **46**, 3077–3107.
- Ebert, E. E., 2008: Fuzzy verification of high-resolution gridded forecasts: A review and a proposed framework. *Meteor. Applic.*, **15**, 51–64.
- Eckhardt, S., and A. Stohl, 2004: A 15-year climatology of warm conveyor belts. *J. Climate*, **17**, 218–237.
- Emanuel, K. A., 1999: Thermodynamic control of hurricane intensity. *Nature*, **401**, 665–669.
- Gemmell, W., B. Katz, and X. Li, 2007: Daily real-time global sea surface temperature – high resolution analysis at NOAA/NCEP. NOAA/NWS/NCEP/MMAB Office Note Number 260, 39 pp.
- Gorbunov, M. E., H. -H. Benzon, A. S. Jensen, M. S. Lohmann, and A. S. Nielsen, 2004: Comparative analysis of radio occultation processing approaches based on Fourier integral operators. *Radio Sci.*, **39**, RS6004, doi:10.1029/2003RS002916.
- Grell, G. A., and D. Devenyi, 2002: A generalized approach to parameterizing convection combining ensemble and data assimilation techniques. *Geophys. Res. Lett.*, **29**, Article 1693.
- Haidvogel, D. B., Arango, H. G., Budgell, W. P., Cornuelle, B. D., Curchitser, E., DiLorenzo, E., Fennel, K., and Co-authors, 2008: Regional ocean forecasting in terrain-following coordinates: model formulation and skill assessment. *J. Comp. Phys.*, **227**, 3595–3624.
- Healy, S. and J.-N. Thépaut, 2006: Assimilation experiments with CHAMP GPS radio occultation measurements. *Quart. J. Roy. Meteor. Soc.*, **132**, 605–623.
- Hong, S.-Y., J. Dudhia, and S.-H. Chen, 2004: A revised approach to ice microphysical processes for the bulk parameterization of clouds and precipitation. *Mon. Wea. Rev.*, **132**, 103–120.
- Hong, S.-Y., and J.-O.J. Lim, 2006: The WRF single-moment 6-class microphysics scheme (WSM6). *J. Korean Meteor. Soc.*, **42**, 129–151.
- Hopkins, E.J., cited 1996: Radiosondes an upper air probe. [Available on-line at <http://www.aos.wisc.edu/~hopkins/wx-inst/wxi-raob.htm>].

- Huang, C.-Y., Y.-H. Kuo, S.-Y. Chen, A.P. Ansetty, and C.-J. Wang, 2007: The assimilation of GPS radio occultation data and its impact on rainfall predictions along the west coast of India during monsoon 2002. *Pure Appl. Geophys.*, **164**, 1577-1591.
- Huang, X.Y., Q. Xiao, D.M. Barker, X. Zhang, J. Michalakes, W. Huang, and Co-authors, 2009: Four-dimensional variational data assimilation for WRF: Formulation and preliminary results. *Mon. Wea. Rev.*, **137**, 299-314.
- Jacobs, N.A., G.M. Lackmann, S. Raman, 2005: 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.
- Janjic, Z. I., 2002: Nonsingular implementation of the Mellor-Yamada Level 2.5 Scheme in the NCEP Meso model, NCEP Office Note, No. 437, 61 pp.
- Jensen, A.S., M.S. Lohmann, H.-H. Benzon, and A.S. Nielsen, 2003: Full spectrum inversion of radio occultation signals. *Radio Sci.*, **38**, 1040, doi:10.1029/2002RS002763.
- Knippertz, P., and H. Wernli, 2009: A Lagrangian climatology of tropical moisture exports into the Northern Hemisphere extratropics. *J. Climate*, **23**, 987-1003.
- Kocin, P.J. and L.W. Uccellini, 2004: *Northeast snowstorms*. Vols. 1 and 2, *Meteor. Monogr.* No. 54. Amer. Met. Soc., 818 pp.
- Kessler, E., 1969: On the distribution and continuity of water substance in atmospheric circulation, *Meteor. Monogr.*, **32**, Amer. Meteor. Soc., 84 pp.
- Kumar, N., G. Voulgaris, and J. C. Warner, 2011: Implementation and modification of a three-dimensional radiation stress formulation for surf zone and rip-current applications, *Coastal Engineering*, **58**, 1097-1117, doi:10.1016/j.coastaleng.2011.06.009.
- Kuo, Y.H., S. Low-Nam, and R.J. Reed, 1991: 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.
- Kuo, Y.-H., W. S. Schreiner, J. Wang, D. L. Rossiter, and Y. Zhang, 2005: Comparison of GPS radio occultation soundings with radiosondes, *Geophys. Res. Lett.*, **32**, L05817, doi:10.1029/2004GL021443.
- LaCasse, K.M., M.E. Splitt, S.M. Lazarus, and W.M. Lapenta, 2008: The impact of high-resolution sea surface temperatures on the simulated nocturnal Florida marine boundary layer. *Mon. Wea. Rev.*, **136**, 1349-1372.
- Lackner, B. C., A. K. Steiner, and G. Kirchengast, 2011: Atmospheric Climate Change Detection by Radio Occultation Data Using a Fingerprinting Method. *J. Climate*, **24**, 5275-5291, doi:10.1175/2011JCLI3966.1.
- Lang, S., W.-K. Tao, R. Cifelli, W. Olson, J. Halverson, S. Rutledge, and J. Simpson, 2007: Improving simulations of convective system from TRMM LBA: Easterly and Westerly regimes. *J. Atmos. Sci.*, **64**, 1141-1164.
- Lang, S., W.-K. Tao, X. Zeng, and Y. Li, 2011: Reducing the Biases in Simulated Radar Reflectivities from a Bulk Microphysics Scheme: Tropical Convective Systems. *J. Atmos. Sci.*, **68**, 2306-2319.
- Larson, J., Jacob, R., Ong, E., 2004. The Model Coupling Toolkit: A New Fortran90 Toolkit for Building Multiphysics Parallel Coupled Models. Preprint ANL/MCSP1208-1204. Mathematics and Computer Science Division, Argonne National Laboratory, 25 pp.

- Lim, K.-S. S., and S.-Y. Hong, 2010: 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.
- Lin, Y.-L., R. D. Farley, and H. D. Orville, 1983: Bulk parameterization of the snow field in a cloud model. *J. Climate Appl. Meteor.*, **22**, 1065–1092.
- Liu, H., J. Anderson, and Y.-H. Kuo, 2012: Improved Analyses and Forecasts of Hurricane Ernesto's Genesis Using Radio Occultation Data in an Ensemble Filter Assimilation System. *Mon. Wea. Rev.*, **140**, 151–166.
- Marzban C, and S. Sandgathe, 2006: Cluster analysis for verification of precipitation fields. *Weather and Forecasting*, **21**, 824–838.
- Mellor, G. L., and T. Yamada, 1982: Development of a turbulence closure model for geophysical fluid problems. *Rev. Geophys. Space Phys.*, **20**, 851–875.
- Ma, Z., Y. -H. Kuo, F. M. Ralph, P. J. Neiman, G. A. Wick, E. Sukovich, and B. Wang, 2011: Assimilation of GPS radio occultation data for an intense atmospheric river with the NCEP regional GSI system. *Mon. Wea. Rev.*, **139**, 2170–2183.
- Mass, C. F., D. Ovens, K. Westrick, and B. A. Colle, 2002: Does increasing horizontal resolution produce more skilled forecasts? *Bull. Amer. Meteor. Soc.*, **89**, 407–430.
- Mellor, G. L., and T. Yamada, 1982: Development of a turbulence closure model for geophysical fluid problems. *Rev. Geophys. Space Phys.*, **20**, 851–875.
- Michalakes, J., S. Chen, J. Dudhia, L. Hart, J. Klemp, J. Middlecoff, and W. Skamarock, 2001: Development of a next generation regional weather research and forecasting model. Argonne National Laboratory Preprint, ANL/MCS-p868-0101, 8pp.
- Milbrandt, J. A., M. K. Yau, 2005: A multimoment bulk microphysics parameterization. Part i: Analysis of the role of the spectral shape parameter. *J. Atmos. Sci.*, **62**, 3051–3064.
- Monin, A.S., and A. M. Obukhov, 1954: Basic laws of turbulent mixing in the surface layer of the atmosphere. *Tr. Akad. Nauk SSSR Geophys. Inst.*, **24**, 163–187.
- Mlawer, E. J., S. J. Taubman, P. D. Brown, M. J. Iacono, and S. A. Clough, 1997: Radiative transfer for inhomogeneous atmosphere: RRTM, a validated correlated-k model for the longwave. *J. Geophys. Res.*, **102**, 16,663–16,682.
- Morrison, H., G. Thompson, and V. Tatarskii, 2009: 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.
- Mote, T. L., D. W. Gamble, S. J. Underwood, M. L. Bentley, 1997: Synoptic-scale features common to heavy snowstorms in the Southeast United States. *Wea. Forecasting*, **12**, five-23.
- Mukai, A. Y., Westerink, J. J., Luettich, R. A., 2001: Guidelines for using the east coast 2001 database of tidal constituents within the Western North Atlantic Ocean, Gulf of Mexico and Caribbean Sea. US Army Corps of Engineers Coastal and Hydraulic Engineering Technical Note (IV-XX).
- NCAR cited 2012a: WRF model version 3.2: Updates. [Available online at <http://www.mmm.ucar.edu/wrf/users/wrfv3.2/updates-3.2.html>].
- NCAR cited 2012b: WRF model version 3.3: Updates. [Available online at <http://www.mmm.ucar.edu/wrf/users/wrfv3.3/updates-3.3.html>].
- NCAR cited 2012c: WRF preprocessing system 3.2: Updates. [Available online at <http://www.mmm.ucar.edu/wrf/users/wpsv3.2/updates-3.2.html>].

- NCAR cited 2012d: WRF preprocessing system 3.3: Updates. [Available online at <http://www.mmm.ucar.edu/wrf/users/wpsv3.3/updates-3.3.html>].
- NCAR cited 2012e: User's guide for the Advanced Research WRF (ARW) modeling system version 3.3. [Available online at http://www.mmm.ucar.edu/wrf/users/docs/user_guide_V3.3/contents.html].
- NCDC, cited 2008: Billion dollar weather disasters 1980-2008. [Available online at: <http://www.ncdc.noaa.gov/img/reports/billion/billion2008.pdf>]
- Nicholls and Decker, 2012: Impact of cycled assimilation of radio occultation data on nor'easter simulations. In-progress.
- Olabarrieta, M., J. C. Warner, and N. Kumar, 2011: Wave-current interaction in Willapa Bay, *J. Geophys. Res.*, **116**, C12014, doi:10.1029/2011JC007387.
- Olabarrieta, M., J. C. Warner, B. Armstrong, J. B. Zamboni, and R. Hei, 2012: Ocean-atmosphere dynamics during Hurricane Ida and Nor'ida: An application of the Coupled Ocean-Atmosphere-Wave-Sediment Transport (COAWST) modeling system. *Ocean Modelling*, **44**, 112-137.
- O'Neill, L.W., D.B. Chelton, and S.K. Esbensen, 2003: Observations of SST-induced perturbations of the wind stress field over the Southern Ocean on seasonal timescales. *J. Climate*, **18**, 2340-2354.
- Park, S., and C. S. Bretherton, 2009: The University of Washington shallow convection and moist turbulence schemes and their impact on climate simulations with the community atmosphere model. *J. Climate*, **22**, 3449-3469.
- Poli, P., S. B. Healy, and D. P. Dee, 2010: Assimilation of Global Positioning System radio occultation data in the ECMWF ERA-Interim reanalysis. *QJRM*, **136**, 1972-1990.
- Pollard, R. T., P. B. Rhines, and R. O. R. Y. Thompson, 1972: The deepening of the wind-mixed layer. *Geophy. Fluid Dyn.*, **4**, 381-404.
- Rabier, F., E. Klinker, P. Courtier, A. Hollingsworth, 1996: Sensitivity of forecast errors to initial conditions. *Quart. J. Roy. Meteor. Soc.*, **122**, 121-150.
- Reisner, J. R., R. M. Rasmussen, and R. T. Bruintjes, 1998: Explicit forecasting of supercooled liquid water in winter storms using the MM5 mesoscale model. *Quart. J. Roy. Meteor. Soc.*, **124**, 1071-1107.
- Ren, X., W. Perrie, Z. Long, and J. Gyakum, 2004: Atmosphere-ocean coupled dynamics of cyclones in the midlatitudes. *Mon. Wea. Rev.*, **132**, 2432-2451.
- Robinson, D. A., 2005: The New Jersey Weather and Climate Network: Providing environmental information for a myriad of applications. Preprints, 15th Conf. on Applied Climatology, Savannah, GA, Amer. Meteor. Soc., J2.1. [Available online at <http://ams.confex.com/ams/pdfpapers/94206.pdf>.]
- Rutledge, S. A., and P. V. Hobbs, 1984: The mesoscale and microscale structure and organization of clouds and precipitation in midlatitude cyclones. XII: A diagnostic modeling study of precipitation development in narrow cloud-frontal rainbands. *J. Atmos. Sci.*, **20**, 2949-2972.
- Serreze, M.C., 1995 : Climatological aspects of cyclone development and decay in the artic. *Atmos.-Ocean*, **33**, 1-23.
- Serreze, M. C., F. Carse, R.G. Barry and J.C. Rogers, 1997: Icelandic low cyclone activity: climatological features, linkages with the NAO and relationships with recent changes in the Northern Hemisphere circulation. *J. Climate*, **10**, 453-464.

- Shchepetkin, A.F., McWilliams, J.C., 2005: The Regional Ocean Modeling System: a split-explicit, free-surface, topography-following coordinates ocean model. *Ocean Modelling*, **9**, 347–404.
- Shi, J. J., W.-K. Tao, T. Matsui, R. Cifelli, A. Hou, S. Lang, A. Tokay, N.-Y. Wang, C. Peters-Lidard, G. Skofronick-Jackson, S. Rutledge, and W. Petersen, 2010: 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.
- Shutts, G. J., 2005: A kinetic energy backscatter algorithm for use in ensemble prediction systems. *Quart. J. Roy. Meteor. Soc.*, **612**, 3079–3102.
- Skamarock, W.C., J. B. Klemp, J. Dudhia, D. O. Gill, D. M. Barker, M. G. Duda, X.-Y. Huang, W. Wang, and J. G. Powers, 2008: A description of the advanced research WRF version 3. NCAR Tech. Note NCAR/TN-475+STR, 125 pp.
- Sokolovskiy, S. 2001: Tracking of tropospheric radio occultation signals from low Earth orbit. *Radio. Sci.*, **36**, 483–498.
- Steiner, A. K., B. C. Lackner, F. Ladstädter, B. Scherllin-Pirscher, U. Foelsche, and G. Kirchengast, 2011: GPS radio occultation for climate monitoring and change detection. *Radio Sci.*, **46**, RS0D24, doi:10.1029/2010RS004614.
- Sutyrin G. G., and A. P. Khain, 1984: Effect of ocean–atmosphere interaction on the intensity of a moving tropical cyclone. *Izv. Acad. Sci. USSR, Atmos. Oceanic Phys.*, **20**, 697–703.
- Tao, W.-K., J.J. Shi, S.S. Chen, S. Lang, P.-L. Lin, S.-Y. Hong, C. Peters-Lidard, and A. Hou: The impact of microphysical schemes on hurricane intensity and track. *Asia-Pacific J. Atmos. Sci.*, **47**, 1–16.
- Thompson, Gregory, Paul R. Field, Roy M. Rasmussen, William 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.
- Tikinaga, H., Y. Tanimoto, and S.P. Xie, 2005: SST-induced surface wind variations over the Brazil-Malvinas confluence: Satellite and in-situ observations. *J. Climate*, **18**, 3470–3482.
- Trenberth, K.E., 1978. On the interpretation of the diagnostic quasi-geostrophic omega equation (for vertical atmospheric motion). *Mon. Wea. Rev.*, **106**, 131–137.
- Trenberth, K.E., and Coauthors, 2007: Observations: Surface and Atmospheric Climate Change. *Climate Change 2007: The Physical Science Basis*, S. Solomon et al., Eds., Cambridge University Press, 235–336.
- United States Census Bureau, cited 2012. New Jersey. [available online at <http://quickfacts.census.gov/qfd/states/34000.html>].
- Velden, S., M. Goodman, and T. Merrill, 1991: Western north pacific tropical cyclone intensity estimation from NOAA polar-orbiting satellite microwave data. *Mon. Wea. Rev.*, **119**, 159–168.
- Wang, W., C. Bruyère, M. Duda, J. Dudhia, D. Gill, H.-C. Lin, J. Michalakes, S. Rizvi, and X. Zhang cited 2012: Weather Research and Forecasting ARW version 3 modeling system users' guide. [Available online at http://www.mmm.ucar.edu/wrf/users/docs/user_guide_V3.3/contents.html].
- Ware, R., and Coauthors, 1996: GPS sounding of the atmosphere from low Earth orbit: Preliminary results. *Bull. Amer. Meteor. Soc.*, **77**, 19–40.

- Warner, J.C., C.R. Sherwood, R. P. Signell, C. Harris, H G. Arango, 2008: Development of a three-dimensional, regional, coupled wave, current, and sediment-transport model. *Computers and Geosciences* 34, 1284–1306.
- Warner, J.C., B. Armstrong, R. He, and J. B. Zambon, 2010: Development of a Coupled Ocean-Atmosphere-Wave-Sediment Transport (COAWST) modeling system. *Ocean Modeling*, **35**, 230-244.
- Weisman, M. L., C. Davis, W. Wang, K. W. Manning and J. B. Klemp, 2008: Experiences with a 0-36-h explicit convective forecasts with the WRF-ARW model. *Wea. Forecasting*, **23**, 407-437.
- Wu, Longtao, Grant W. Petty, 2010: Intercomparison of Bulk Microphysics Schemes in Model Simulations of Polar Lows. *Mon. Wea. Rev.*, **138**, 2211–2228.
- Yao, Y., W. Pierre, W. Zhang, J. Jiang, 2008: Characteristics of atmosphere-ocean interactions along North Atlantic extratropical storm tracks. *J. Geophys. Res.*, **113**, doi:10.1029/2007JD008854.
- Zhang, D.-L., and R.A. Anthes, 1982: A high-resolution model of the planetary boundary layer—sensitivity tests and comparisons with SESAME–79 data. *J. Appl. Meteor.*, **21**, 1594–1609.