A COMPARISON OF POLAR WRF AND POLAR MM5 IN ANTARCTIC SURFACE FORECASTS

Jordan G. Powers and Kevin W. Manning

Mesoscale and Microscale Meteorology Division
Earth and Sun Systems Laboratory, National Center for Atmospheric Research
Boulder, Colorado, USA

1. BACKGROUND

The Weather Research and Forecasting (WRF) model has been running in the Antarctic Mesoscale Prediction System (AMPS) (Powers et al. 2003) since late 2005. AMPS is a real-time, experimental mesoscale modeling system tailored for Antarctica and providing high-resolution NWP guidance to the Unites States Antarctic Program as well as number of other international efforts. Within AMPS the Advanced Research WRF (ARW) (Skamarock et al. 2005) runs alongside a polar-modified version of its predecessor, the MM5 (Grell et al. 1995), although an aim has been to settle on a single model.

The MM5 has had a polar-modified version for a number of years ("polar MM5"; see, e.g., Cassano et al. 2001), while polar development of the newer ARW ("polar WRF") has been much more recent ((Bromwich and Hines 2006). The goal of the development of polar WRF is to improve the model's capabilities for high-latitude research and forecasting. The testing and evaluation of polar WRF began last year (Powers and Manning 2007), and this version began running in AMPS in August 2007. USAP forecaster experience with polar WRF in AMPS, and its performance relative to polar MM5, for the 2007–2008 field season have both been positive (Space and Naval Warfare Research Center (USAP forecasters), personal communications). A more quantitative analysis is needed, however, to compare polar MM5 (PMM5) and polar WRF (PWRF) in AMPS. This study thus examines both models running in AMPS over Antarctica, with an eye to determining whether PWRF is now comparable to PMM5. Two test periods are considered and forecast error statistics compared. The forecast parameters considered here are of surface conditions: temperature and wind speed. Statistical significance testing is performed to distinguish the differences in forecast skill.

2. POLAR WRF AND EXPERIMENTS

The polar modifications to WRF (ARW) primarily involve adjusted lower boundary and land surface characteristics. First, PWRF accounts for fractional sea ice coverage in grid cells. Grid cells are quantified as fractionally covered if not fully ice-covered or fully ice-free, as indicated by daily sea ice analyses. In contrast, in the default ARW grid cells are either fully open or fully covered.

Many of the polar changes are to the Noah land surface model (LSM). These are: the use of the latent heat of sublimation for calculations of latent heat fluxes over ice surfaces (permanent ice, sea ice, and snow cover); the assumption of ice saturation when calculating surface saturation mixing ratios over ice; an increase in the value of snow albedo; adjustment of snow density; adjustment of snow heat capacity and thermal diffusivity for the subsurface layers; an increase in the emissivity value for snow; and a modification of the skin temperature calculation (enforcing a sensible heat/latent heat/radiation energy balance at the surface). Other modifications outside of the Noah LSM are: a stability-dependent formulation for the mixing length, and a modified initialization of low-level air temperatures based on an interpolation of the previous forecast's subsurface temperature and the first-guess field's air temperatures at its lowest levels. In addition, soil temperatures are cycled in polar WRF; that is, the previous forecasts' subsurface temperatures are used in new forecast initializations. Lastly, the shortwave radiation scattering parameter is reduced in polar WRF to account for less clear-air scattering.

The two periods evaluated are 26 January–15 February 2007 and 6–23 April 2007, which will be referred to as the January and April periods, respectively. These periods allow examination of both warm and cold season performance. Both WRF and MM5 were run over 5 domains, ranging in grid size from 60 km to 2.2 km, shown in Fig. 1. The grids are referred to as follows: 60-km outer grid— Domain 1; 20-km Antarctic grid— Domain 2; 6.7-km western Ross Sea grid— Domain 3; 6.7-km South Pole grid— Domain 4; 2.2-km Ross Island grid— Domain 5. The forecast lengths were 48 hours on all grids. First-guess fields and boundary conditions were derived from the GFS (Global Forecasting System) global model analyses and forecasts. The first-guess fields were reanalyzed with observations (surface, upper-air, satellite) using the 3-dimensional variational data assimilation system, WRF-Var.

The variables verified are temperature (T) and wind speed (WSP) at the lowest model half-level (vertical level) (about 13 m AGL) and 2-m temperature (T2). Surface AWS (automatic weather station) data are used for the verification. The number of AWS stations available within each domain for verification varies with the given month and variable, with Domain 2 reporting 22–28, Domain 3 reporting 15–18, Domain 4 reporting

3, and Domain 5 reporting 12–14. Model bias (mean error— ME), mean absolute error (MAE), and root mean square error (RMSE) are calculated for forecast hours 12, 24, 36, and 48. The analysis includes significance testing to determine if differences in error means between models are statistically significant. Student's t-tests are applied, and all significance is evaluated at the 95% confidence level.

3. RESULTS

a. Significance testing

For WRF and MM5 for a given period, domain, forecast hour, and variable, differences in the experiment mean statistics (bias, MAE, RMSE) were analyzed to determine statistical significance. In addition, one-tailed tests were applied to determine which experiment had the lower MAE and RMSE, i.e., which model performed better. It is first seen that for most of the variables, times, etc., the differences in mean errors are not statistically significant. For example, for both periods, all four forecast hours, all four domains, and all three variables (T, WSP, T2), there are 96 permutations and possible model comparisons. The differences in the RMSEs for WRF and MM5 were not statistically significant for 73 (76.0%) of these. For 12 (12.5%) of these PWRF had a statistically significantly lower RMSE, while for 11 (11.5%) of these the MM5 had the lower error. For MAE, the total was 77 (80.2%) no significant difference, 9 (9.4%) WRF lower, and 10 (10.4%) MM5 lower. Considering just the variables of (model) surface temperature and wind speed (64 comparisons), the RMSE results were 47 (73.4%) no significant differences, 9 (14.1%) WRF better, and 8 (12.5%) MM5 better. MAE results were about the same, with 51 (79.7%) no difference, 6 (9.4%) WRF better, and 7 (10.9%) MM5 better. Thus, the large majority of comparisons show that polar WRF is now comparable to polar MM5. Furthermore, where there are significant differences in the mean values of the error statistics, the results are about balanced in terms of either model having an edge.

Differences in the mean values of biases are found to be statistically significant at a somewhat higher frequency. For the 96 comparisons of T, WSP, and T2, the differences in mean bias are significant at the given level for 56 (58.3%) comparisons and not significant for 40 (41.7%) of the comparisons. For the 64 comparisons of T and WSP, the numbers are 33 (51.6%) significant and 31 (48.4%) not significant. Note, however, that where instances of significant difference of the two mean biases (of T and T2) involve one being positive and one negative, there are cases where neither bias is statistically significantly different from 0, i.e., the departure from 0 for each model is not significant. In Tab. 1, for example (discussed below), the hour 12 and hour 48 mean T biases for Domain 2 in April are not significantly different from 0, although they are significantly different from each other

b. Error time series

While domain-averaged errors have been calculated from verifications at AWS's across each domain, the average errors for individual sites through the 48-hour forecasts have been determined also. Figures 2–7 present plotted station examples from different domains and regions.

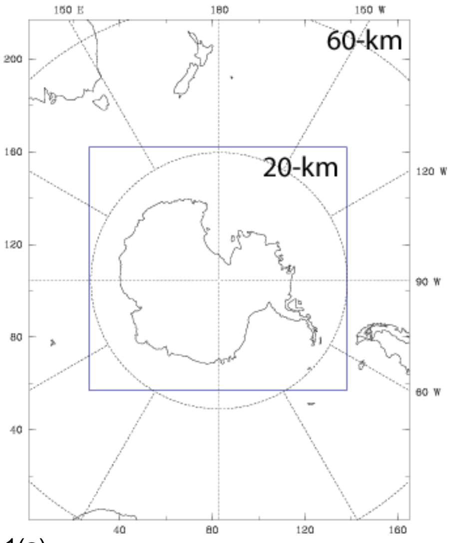
Figures 2(a)–(d) show surface temperature (T) bias (solid), MAE (dotted)[1], and RMSE (dashed) for South Pole for WRF and MM5 for both periods. The South Pole forecasts reflect the 6.7-km domain 4 (Fig. 1(b)). These plots also show the average values of the errors for the whole forecast period. For example, the WRF and MM5 average biases for the January period (2(a),(b)) are 2.6°C and 5.1°C, respectively, while for April (2(c),(d)) they are .55°C for WRF and 8.1°C for MM5. As can be seen, the bias for the MM5 is very high (>10C) during the first 12 hours (off the scale) and is higher than WRF's. The overall MAEs for WRF are lower for both months— 3.3°C (WRF) v. 5.2°C (MM5) for January and 3.0°C (WRF) and 8.4°C (MM5) for April.

Figure 3 shows the January results for Dome A, on the Antarctic plateau, within Domain 2 (locations of sites indicated in figure insets; see captions). For January (3(a),(b)), WRF's bias is negligible (.04°C), while that for MM5 is 1.4°C, and the MAE and RMSE are correspondingly lower for WRF. For April (not shown), WRF's bias (.96°C), MAE (4.9°C), and RMSE (5.8°C) are lower than those of the MM5 (3.0°C, 5.4°C, and 6.4°C, respectively).

For a look at performance in the critical (to the USAP) Ross Island region, Figs. 4 and 5 show results for Pegasus North and Minna Bluff. These are within the 2.2-km Ross Island grid, Domain 5 (Fig. 1(b)). For Pegasus North these reveal overall negative (cold) temperature biases for both models for January and positive (warm) biases for April. Temperature error magnitudes are slightly lower in MM5 for the warm season (January) period, but lower for WRF in the cold season (April) period. For Minna Bluff, January (not shown) biases are positive for MM5 (.95°C), but negative for WRF (-.99°C), and the MAE and RMSE are comparable. For April (Figs. 5(a),(b)), WRF displays a lower positive bias than the MM5 (1.6°C v. 3.0°C), and lower MAE and RMSE. In the context of the development of polar WRF and its eventual replacement of the polar MM5 in AMPS, all of the temperature results are encouraging in what is *not* seen: there is no significant (e.g., $o(5 – 10)$ °C) warm bias over the continent seen in PWRF. This is in contrast to results looking at a previous version of PWRF in more limited AMPS testing described by Powers and Manning (2007).

---

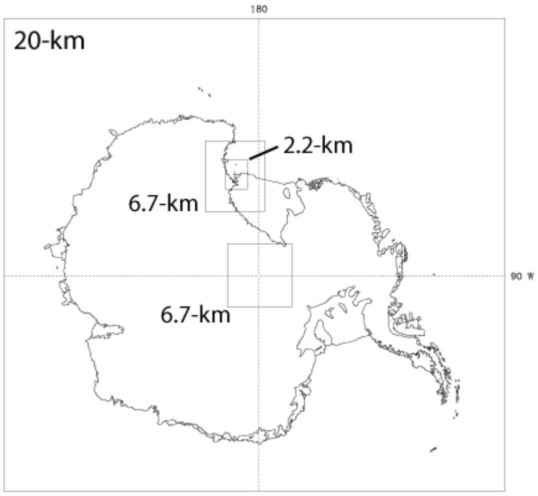
[1] Depending on the quality of the hardcopy reproduction of the figures, the MAE curves may be indistinct.



1(a)

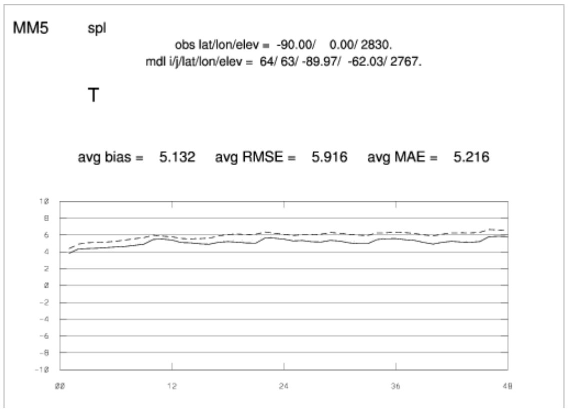
Figure 6 offers the January wind speed results for South Pole. For both periods the errors are similar and low overall. Slight positive biases are seen for both WRF and MM5, virtually identical for January (~.6 ms^{-1} for both), and WRF about .3 ms^{-1} higher for April (.9 ms^{-1} v. 1.2 ms^{-1}) (not shown). MAEs are less than 2.0 ms^{-1} for both models for both periods.

Lastly, Fig. 7 provides a comparison of wind speed errors at Pegasus North. For January (7(a),(b)), biases are positive for both WRF (.87 ms^{-1}) and MM5 (1.5 ms^{-1}), with slightly lower MAE and RMSE for WRF. The results are similar for April (not shown), although both models have a greater overall positive bias, 2.9 ms^{-1} for WRF and 3.1 ms^{-1} for MM5.

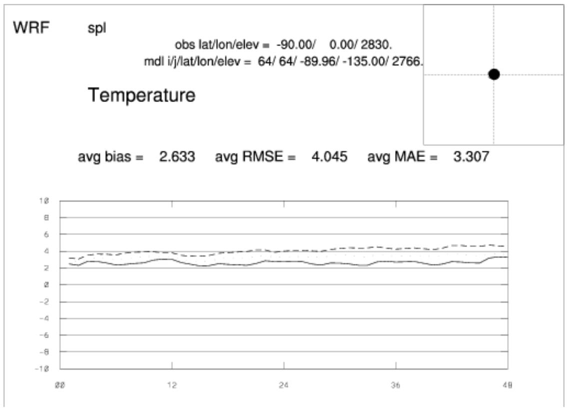


1(b)

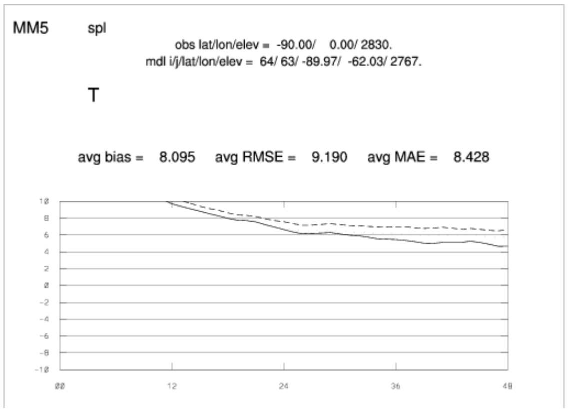
Fig. 1: Model grids. (a) 60-km (Domain 1) and 20-km (Domain 2) grids. (b) 20-km, 6.7-km (Domain 3—western Ross Sea; Domain 4— South Pole), and 2.2-km (Domain 5 —Ross Island) grids.



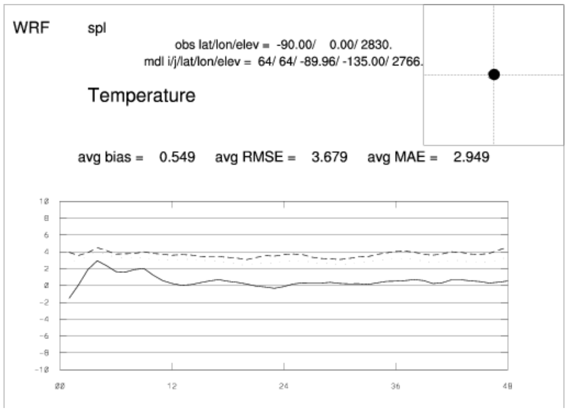
2(a)



2(b)

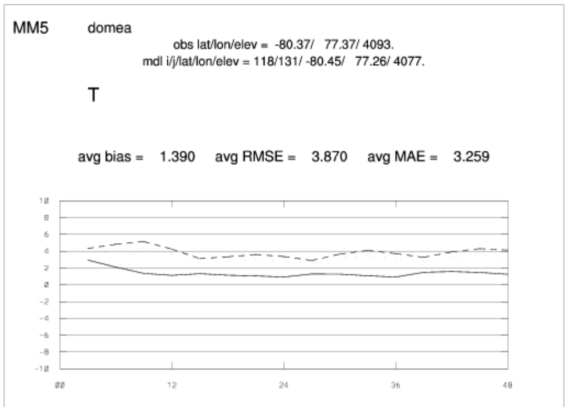


2(c)

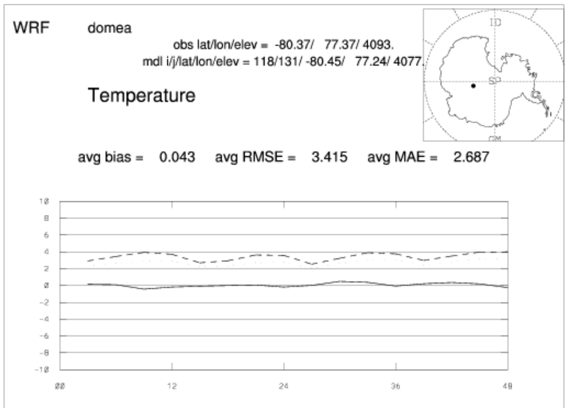


2(d)

Fig. 2: South Pole surface temperature bias, RMSE, and MAE (°C). Jan.= Jan.-Feb. 2007 test period. Apr.= April 2007 test period. (a) MM5 Jan. (b) WRF Jan. (c) MM5 Apr. (d) WRF Apr.

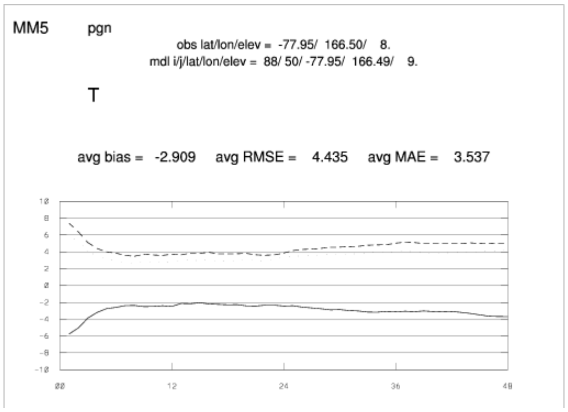


3(a)

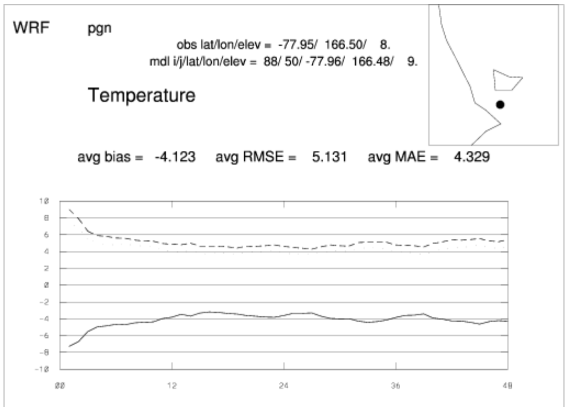


3(b)

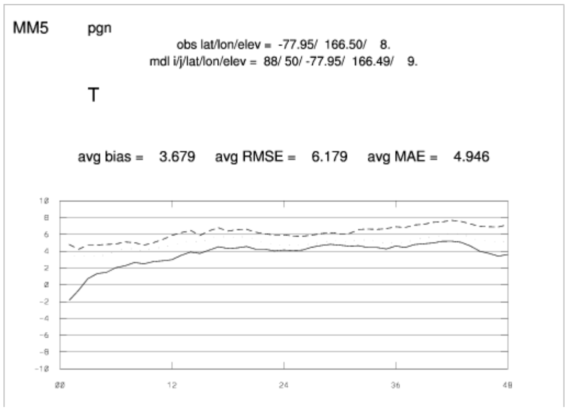
Fig. 3: Dome A January surface temperature bias, RMSE, and MAE (°C). Location of Dome A in Domain 2 shown in inset in (b). (a) MM5 Jan. (b) WRF Jan.



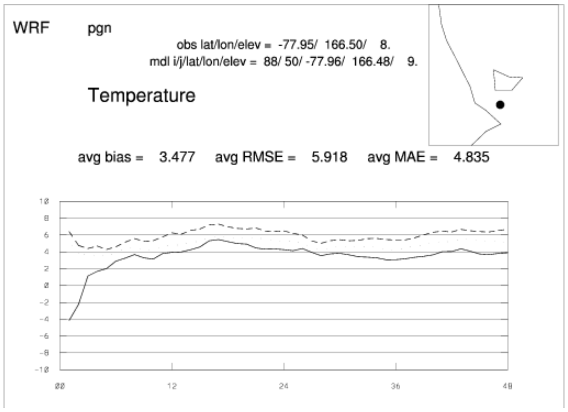
4(a)



4(b)

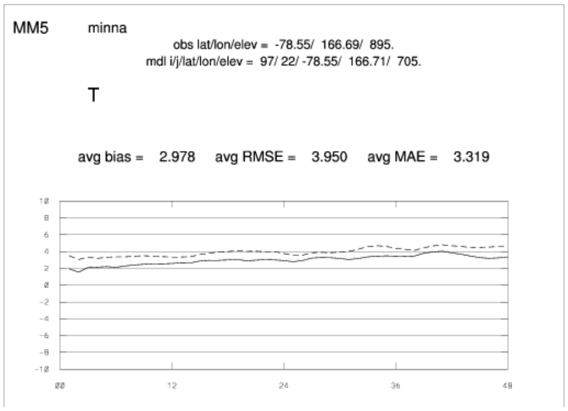


4(c)

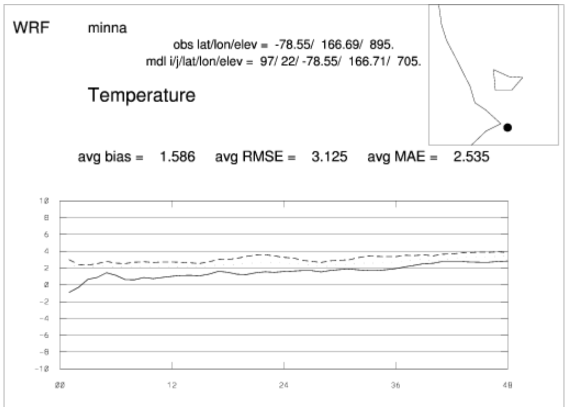


4(d)

Fig. 4: Pegasus North surface temperature bias, RMSE, and MAE (°C). Location of Pegasus North in Domain 3 shown in inset in (b), (d). (a) MM5 Jan. (b) WRF Jan. (c) MM5 Apr. (d) WRF Apr.

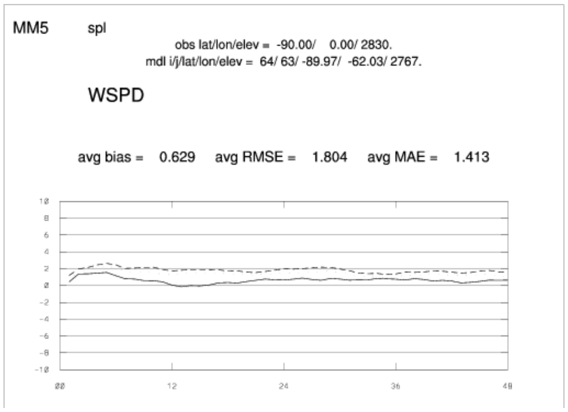


5(a)

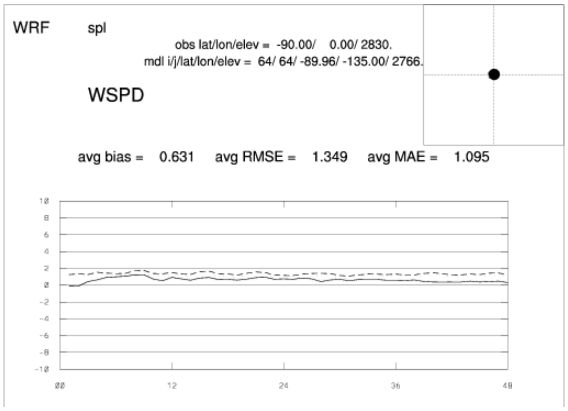


5(b)

Fig. 5: Minna Bluff surface temperature bias, RMSE, and MAE (°C). Location of Minna Bluff in Domain 3 shown in inset in (b). (a) MM5 Apr. (a) WRF Apr.

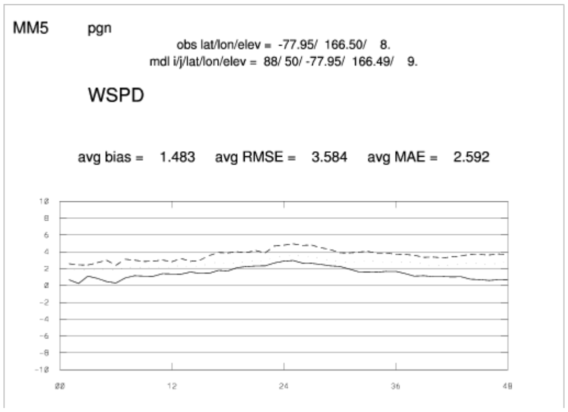


6(a)

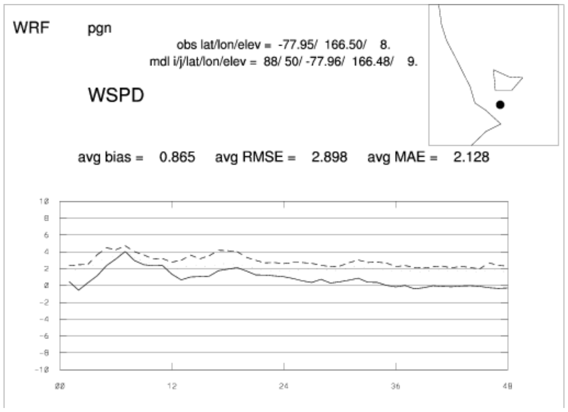


6(b)

Fig. 6: South Pole surface wind speed bias, RMSE, and MAE (ms^{-1}). (a) MM5 Jan. (b) WRF Jan.



7(a)



7(b)

Fig. 7: Pegasus North surface wind speed bias, RMSE, and MAE (ms^{-1}). Location of Pegasus North in Domain 3 shown in inset in (b). (a) MM5 Jan. (b) WRF Jan.

Domain 2

January Surface Temperature (°C)

	Bias			RMSE			MAE		
	WRF	MM5	∆	WRF	MM5	∆	WRF	MM5	∆
HR 12	-1.73	**-1.24**	-0.50	3.36	3.48	-0.12	2.85	2.91	-0.06
HR 24	-1.70	-1.58	-0.12	3.35	3.59	-0.24	2.82	2.97	-0.15
HR 36	-1.67	-1.71	0.04	**3.50**	3.84	-0.34	**2.93**	3.23	-0.29
HR 48	- 1.66	-1.66	0.00	**3.55**	3.85	-0.30	**2.96**	3.20	-0.24

April Surface Temperature (°C)

	Bias			RMSE			MAE		
	WRF	MM5	∆	WRF	MM5	∆	WRF	MM5	∆
HR 12	-0.77	0.61	-1.39	4.24	4.79	-0.56	3.47	4.02	-0.55
HR 24	-0.36	-0.22	-0.14	4.50	4.90	-0.39	3.69	4.03	-0.34
HR 36	0.18	-0.32	0.50	4.67	4.91	-0.24	3.74	4.07	-0.33
HR 48	0.63	-0.52	1.14	4.96	5.07	-0.10	3.96	4.16	-0.20

January Surface Wind Speed (ms^{-1})

	Bias			RMSE			MAE		
	WRF	MM5	∆	WRF	MM5	∆	WRF	MM5	∆
HR 12	2.18	**1.27**	0.91	3.30	3.20	0.10	2.79	2.62	0.17
HR 24	1.94	**1.52**	0.42	3.31	3.56	-0.25	2.75	2.85	-0.10
HR 36	1.75	1.44	0.31	3.33	3.48	-0.15	2.75	2.82	-0.07
HR 48	1.81	**1.27**	0.53	3.43	3.50	-0.07	2.82	2.79	0.02

April Surface Wind Speed (ms^{-1})

	Bias			RMSE			MAE		
	WRF	MM5	∆	WRF	MM5	∆	WRF	MM5	∆
HR 12	4.07	**2.20**	1.87	5.61	**4.79**	0.81	5.00	**4.05**	0.95
HR 24	3.76	**3.07**	0.69	5.48	5.26	0.22	4.83	4.56	0.27
HR 36	3.21	2.89	0.32	5.21	5.03	0.18	4.50	4.35	0.15
HR 48	2.95	2.82	0.14	5.21	5.09	0.12	4.35	4.29	0.06

Table 1: Bias (ME), RMSE, and MAE statistics for Domain 2 for surface temperature and wind speed verifications for January and April periods. ∆= error difference of WRF–MM5. Boldface values indicate that given model error is lower at 95% level of statistical significance.

c. Example of domain-averaged statistics

Table 1 presents the surface temperature and wind speed bias (ME), RMSE, and MAE scores for verifications over the continental 20-km grid, Domain 2. The differences of the raw errors appear in the column headed "∆". Where these differences are statistically significant, the model with the lower error is indicated in boldface. One can see the magnitude of the statistically significant difference (∆) to gauge whether it is practically significant.

Overall, the statistically significant differences are few, scattered, and fairly balanced in number. With respect to T, April shows no statistically significant differences in any metric, with error differences in the polar WRF and polar MM5 less than .5 °C. For January, both models show a warm bias of generally less than 2 °C. The MM5 has a lower bias for hour 12 (.5 °C), although WRF has a lower RMSE and MAE for hours 36 and 48, averaging .29 °C less.

For wind speeds, both models have a positive bias for all hours, i.e., they are slightly overforecasting wind speeds. The MM5 shows a statistically significantly lower bias in January for hours 12, 24, and 48; for April its bias is lower for hours 12 and 24. However, the average difference is less than .9 ms^{-1}. The verifications and significance testing for Domains 3, 4, and 5 (tables not presented) yield similar results: error differences do not consistently favor either model, and where statistically significant differences appear, their magnitudes are small.

4. SUMMARY AND CONCLUSIONS

Polar WRF and Polar MM5 have been run in warm and cool season periods over Antarctica and forecast error statistics calculated and compared. 48-hour forecasts were made over the AMPS domains, with verifications done at AWS sites on Domain 2 (20 km continent-wide), Domain 3 (6.7-km western Ross Sea), Domain 4 (6.7-km South Pole), and Domain 5 (2.2-km Ross Island). Test periods of approximately three weeks each were in January–February 2007 and April 2007. Surface temperature and wind speed and 2-m temperature biases, RMSEs, and MAEs were compared.

Overall, polar WRF is found to be comparable to the polar MM5. For the majority (75– 80%) of error comparisons, the differences in metrics are not statistically significant (at the 95% level). For comparisons of surface temperature and wind speed RMSEs and MAEs which did reveal statistically significant differences (i.e., approx. 20% of the error comparisons), the frequency of lower model errors in either PWRF or PMM5 were evenly split.

Plots of error time series for individual AWS sites illustrate the overall comparability of the polar-modified models. Bias and RMSE time series reveal little systematic differences between the levels of errors. Where one model may have somewhat lower error scores in one season at a given site, the other prevails in the other season. It is noteworthy that in a prior examination of the MM5 with WRF in AMPS domains over Antarctica by the authors, WRF was found to have a significant warm bias compared to the MM5 (Powers and Manning 2007). This was true for both an unmodified WRF and for a preliminary version of polar WRF.

Analyses of domain-wide errors for surface temperature and wind speed also reveal the performance of PWRF and PMM5 to be comparable. Where mean error differences are found to be statistically significant, they are roughly split between PWRF and PMM5 being better, and the magnitudes of these statistical differences are not large in practical terms.

The forecast errors considered here have all been for parameters in the surface layer. Upper-air verification work is in progress. In examining the impacts of polar modifications among WRF runs (i.e., WRF v. polar WRF), Powers and Manning (2007) found that these were not significant above the PBL.

The development and evaluation of polar WRF is ongoing, but comparisons with polar MM5 in AMPS for operational purposes will be completed in 2008. Ultimately, the polar modifications for the ARW will be incorporated into the official WRF repository and provided to the WRF community in a future release.

ACKNOWLEDGMENTS

The authors thank Dr. Keith Hines and Le-Sheng Bai of the Polar Meteorology Group, Byrd Polar Research Center, The Ohio State University for their development and contribution of the polar modifications to WRF. The authors thank NSF and the NSF Office of Polar Programs for their support of AMPS.

The National Center for Atmospheric Research is sponsored by the National Science Foundation. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

Bromwich, D.H., and K.M. Hines, 2006: Development and testing of Polar WRF. *Antarctic Meteorological Observation, Modeling, and Forecasting Workshop*, Boulder, CO, Nat'l Center for Atmos. Res.

Cassano, J.J., J.E. Box, D.H. Bromwich, L. Li, and K. Steffen, 2001: Evaluation of Polar MM5 simulations of Greenland's atmospheric circulation. *J. Geophys. Res.*, **106**, 33867–33890.

Grell, G.A., J. Dudhia, and D.R. Stauffer, 1995: A description of the fifth-generation Penn State/NCAR Mesocsale Model (MM5). NCAR Tech. Note TN-398+STR, 122 pp. [Available from UCAR Communications, P.O. Box 3000, Boulder, CO 80307]

Powers, J.G., A.J. Monaghan, A.M. Cayette, D.H. Bromwich, Y.-H. Kuo, and K.W. Manning, 2003: Real-time mesoscale modeling over Antarctica: The Antarctic Mesoscale Prediction System (AMPS). *Bull Amer. Meteor. Soc.*, **84**, 1533–1546.

Powers, J.G., and K.W. Manning, 2007: Polar WRF testing in Antarctica. *8th WRF Users' Workshop*, National Center for Atmos. Res., Boulder, CO.

Skamarock, W.C., J.B. Klemp, J. Dudhia, D.O. Gill, D.M. Barker, W. Wang, and J.G. Powers, 2005: *A description of the Advanced Research WRF, Version 2*. NCAR Tech. Note., NCAR/TN-468+STR, 88 pp. [Available from UCAR Communications, P.O. Box 3000, Boulder, CO 80307]