

## Comprehensive evaluation of polar weather research and forecasting model performance in the Antarctic

David H. Bromwich,<sup>1,2</sup> Francis O. Otieno,<sup>1</sup> Keith M. Hines,<sup>1</sup> Kevin W. Manning,<sup>3</sup> and Elad Shilo<sup>4</sup>

Received 22 May 2012; revised 7 November 2012; accepted 8 November 2012; published 17 January 2013.

[1] Recent versions of the Polar Weather Research and Forecasting model are evaluated over the Antarctic to assess the impact of model improvements, resolution, large-scale circulation variability, and uncertainty in initial and lateral boundary conditions. The model skill differs more between forecasts using different sources of lateral boundary data than between forecasts from different model versions or simulated years. Using the ERA-Interim reanalysis for initial and lateral boundary conditions produces the best skill. The forecasts have a cold summer and a warm winter bias in 2 m air temperatures, with similar but smaller bias in dew point temperatures. Upper air temperature biases are small and remain less than 1 °C except at the tropopause in summer. Geopotential height biases increase with height in both seasons. Deficient downward longwave radiation in all seasons and an under representation of clouds enhance radiative loss, leading to the cold summer bias. Excess summer surface incident shortwave radiation plays a secondary role, because 80% of it is reflected, leading to greater skill for clear compared with cloudy skies. The positive wind speed bias produces a warm surface bias in winter resulting from anomalously large downward flux of sensible heat toward the surface. Low temperatures on the continent limit sublimation and hence the precipitable water amounts over the ice sheet. ERA-Interim experiments with higher precipitable water showed reduced biases in downwelling shortwave and longwave radiation. Increasing horizontal resolution from 60 to 15 km improves the skill of surface wind forecasts.

**Citation:** Bromwich, D. H., F. O. Otieno, K. M. Hines, K. W. Manning, and E. Shilo (2013), Comprehensive evaluation of polar weather research and forecasting model performance in the Antarctic, *J. Geophys. Res. Atmos.*, 118, 274–292, doi:10.1029/2012JD018139.

### 1. Introduction

[2] As a result of advances in modeling, observational data coverage, and communication systems, critical forecasts in support of the U.S. Antarctic Program (USAP) are now routinely made using the polar-optimized version of the Weather Research and Forecasting model (Polar WRF) via the Antarctic Mesoscale Prediction System (AMPS) [Powers *et al.*, 2012]. AMPS forecasts with Polar WRF version 3.0.1 began in August 2008, and the model was used with few configuration changes until April 2011, when an upgrade to Polar WRF 3.2.1 was implemented. Not only did the AMPS model change

but, as computer capability has increased, also has the resolution used by AMPS, which now stands at 1.67 km around McMurdo Station, where the USAP hub is located. However, the WRF model was originally developed for the Northern Hemisphere midlatitudes, and for polar regions WRF has been tested primarily in Arctic environments [Bromwich *et al.*, 2009; Hines and Bromwich, 2008; Hines *et al.*, 2011; Cassano *et al.*, 2011]. Therefore, it is necessary for advancing Antarctic modeling to benchmark its performance in Antarctica.

[3] Relying on these AMPS forecasts, six ski-equipped LC-130 Hercules aircraft flown by the New York Air National Guard completed 359 missions while the 62nd and 446th Airlift Wing flew 72 missions in support of Operation Deep Freeze during the 2011–2012 field season; together they transported at least 6600 people and 5.8 million kilograms of equipment [Hannon, 2012]. In addition to transport logistics, accurate forecasts are also needed for safety on the ground. A number of people have had to be evacuated at various times from the continent as recently as August 2012 [National Science Foundation (NSF), 2012], and tourist visits now exceed 30,000 annually [International Association of Antarctic Tour Operators (IAATO), 2010]. Accurate forecasts will therefore continue to be required for the safety of those working in Antarctica during severe wind storms

<sup>1</sup>Polar Meteorology Group, Byrd Polar Research Center, The Ohio State University, Columbus, Ohio, USA.

<sup>2</sup>Atmospheric Sciences Program, Department of Geography, The Ohio State University, Columbus, Ohio, USA.

<sup>3</sup>National Center for Atmospheric Research, Boulder, Colorado, USA.

<sup>4</sup>Israeli Meteorological Service, Bet Dagan, Israel.

Corresponding author: D. H. Bromwich, Polar Meteorology Group, Byrd Polar Research Center, The Ohio State University, 1090 Carmack Road, Columbus, OH 43210, USA. (bromwich.1@osu.edu45)

©2012. American Geophysical Union. All Rights Reserved.  
2169-897X/13/2012JD018139

and in potential large-scale evacuations in the future. In addition, archived AMPS forecasts have been used extensively to study Antarctic weather and climate [e.g., *Monaghan et al.*, 2003; 2005; *Fogt and Bromwich*, 2008; *Powers*, 2007; *Powers et al.*, 2010, 2012; *Steinhoff*, 2011; *Bromwich et al.*, 2011a; *Nigro et al.*, 2012; *Seefeldt and Cassano*, 2012; *Steinhoff et al.*, 2012].

[4] Model simulations produce data that in the absence of long-term observations can be used to diagnose climatic processes such as those depicted in Antarctic ice cores covering eight glacial cycles [*European Project for Ice Coring in Antarctica (EPICA)*, 2004]. Such applications will require climate models that accurately predict the short-term weather (forecasts) before they can be used to diagnose long-term climate process such as the accelerated rate of ice loss from the continent [*Velicogna*, 2009; *Rignot et al.*, 2011]. Obtaining a complete understanding of Antarctic teleconnections and climate variability using models is an area of active research, and a recent report of the *National Research Council (NRC)* [2011] supports the development of a new generation of robust earth system models.

[5] The skill of Polar WRF already demonstrated in the Arctic may not prevail in Antarctica because of substantial differences between the two polar regions. First, the atmospheric circulation in Antarctica is strongly influenced by the presence of the large ice sheet, with elevations exceeding 3 km in most areas. Any mesoscale model used here must accurately represent the strong katabatic winds that are governed by the balance of gravity, thermal stability, and synoptic forcing [e.g., *Bromwich and Liu*, 1996; *Cassano and Parish*, 2000; *Parish and Bromwich*, 2007]. The cold temperatures limit evaporation and sublimation and make ice the dominant phase of water in Antarctic clouds. High surface albedo [*Laine*, 2008] and relatively pristine skies (low amounts of aerosol) also result in different radiation-atmosphere interactions. Surrounded by a large ocean, Antarctica has limited observational data needed to initialize mesoscale models and to validate forecasts, making the verification of models more difficult than in the Arctic.

[6] Development of the standard WRF is a community effort led by the National Center for Atmospheric Research (NCAR), among others. Benefits of community involvement in the development include rapid and continuous refinements to both the standard and the Polar WRF. The primary drawback to such rapid developments is that newer model versions are subject to fewer evaluations than their predecessors. All past verifications cannot be repeated every time a new model is released. Incompatibilities in the physics parameterizations and other errors (bugs) also arise occasionally from the model developments themselves. In a recent Arctic-wide evaluation, Polar WRF 3.1.1 demonstrated significant Arctic forecast skill. This study expands the evaluations to all of Antarctica on an annual time scale and is the sixth in a series [*Hines and Bromwich*, 2008; *Bromwich et al.*, 2009; *Hines et al.*, 2011; *Wilson et al.*, 2011, 2012] of verification studies documenting ongoing Polar WRF developments. We focus on three factors that influence forecast skill: model improvements, uncertainty in driving data, and interannual variations in the large-scale atmospheric circulation. The impact of model improvements is assessed using four recent versions of Polar WRF (releases 3.0.1, 3.1.1, 3.2.1, and 3.3.1). Bug-fixed subversions are chosen because they include

corrections for problems identified by the user community since the initial release. Improvements that only make the model easy to use, enhance computational efficiency, or speed up diagnosis of model output are not considered.

[7] This study examines the impact on forecast skill of uncertainty in the initial and lateral boundary conditions by comparing Polar WRF forecasts made using the reanalysis from the European Centre for Medium-Range Weather Forecasts (ECMWF; ERA-Interim) [*Dee et al.*, 2011] and the final analysis from the National Centers for Environmental Prediction (NCEP) operational global analysis (GFS-FNL). The relative accuracy of contemporary reanalyses has been examined previously in relation to predicted precipitation, and ERA-Interim fared the best [*Nicolas and Bromwich*, 2011; *Bromwich et al.*, 2011b]. In addition, from a comparison of five reanalyses with Antarctic station observations [*Bracegirdle and Marshall*, 2012], ERA-Interim was found to be the most reliable for mean sea level pressure and 500 hPa height trends. ERA-Interim represents a major advance over previous ECMWF reanalyses because of increased resolution (T255), a more advanced atmospheric model, 4D-Var assimilation system, and the enormous volume of assimilated data. Assimilation of satellite radiance data is especially critical over the Southern Ocean, where there are few synoptic stations. However, reanalyses such as ERA-Interim have the advantage of a fixed modeling system and input data that are usually unavailable in operational forecasting. Thus, many centers rely on operational analyses for lateral boundary conditions. For this reason, this study compares Polar WRF skill when driven by an analysis and a reanalysis. The two cover a reasonable range that can be associated with the uncertainty in the lateral boundary data and are considered sufficient for the objectives here. However, we note that many other sources of lateral and initial conditions exist.

[8] Influence of interannual variations in the large-scale atmospheric circulation on model skill is evaluated by comparing forecast statistics from two different years (1993 and 2007); both made using the same analysis (ERA-Interim). We choose 1993 to allow for a comparison between the statistics given here with those of an earlier Polar MM5 Antarctic verification [*Guo et al.*, 2003], and using 2007 allows us to take advantage of recent advances in the observing network. Strictly speaking, the quality of a reanalysis depends on the number of assimilated observations, and these did differ between 1993 and 2007, thereby compounding differences resulting from interannual variations with those resulting from analysis quality. However, we assume that ERA-Interim representations of Antarctic atmospheric circulation are among the best possible for these years [*Bromwich et al.*, 2011b].

## 2. Model, Experiments, Lower Boundary Conditions, and Observations

### 2.1. Model Description

[9] Polar WRF [*Hines and Bromwich*, 2008; *Hines et al.*, 2011] was developed by the Polar Meteorology Group at The Ohio State University and builds on previous success with an earlier mesoscale model (Polar MM5) [*Bromwich et al.*, 2001] in the Arctic. The overall difference is that Polar WRF is modified with enhancements that adapt the standard

**Table 1.** The WRF Configuration and Physics Options Used

Model Version/Physics	3.0.1	3.1.1	3.2.1
Longwave			
RRTM	X	X	
RRTMG			X
Shortwave			
Goddard	X	X	
RRTMG			X
Microphysics	WRF Single moment 5 class		
Lateral boundary	NCEP Global Final analysis GFS-FNL; 2007)		
Planetary boundary layer	Mellor Yamada-Janjic (MYJ)		
Sea surface temperature (SST)	NCEP RTG (2007)		
Land surface	Noah Land Surface Model (LSM)		
Surface layer	Monin-Obukhov (Janjic-eta)		
Time step	120 s		
Horizontal resolution	60 km		
Cumulus parameterization	Grell-Devenyi		
Model top	10 hPa with damping in the top four levels (8 km depth)		
Latitude/Longitude	121 × 121		
Relaxation zone	10 grid points		
Vertical resolution	39 eta levels		
Sea ice	NSIDC 25 km fractional sea ice		
Integration	48 h forecasts from global analysis		
Spin up	First 24 h used for model spin up		
Base state temperature	273.16 K		

code to the polar regions. These include freezing point of seawater set at 271.36 K, surface roughness of 0.001 m over sea ice and permanent land ice, snow emissivity of 0.98, snow density over sea ice of  $300 \text{ kg m}^{-3}$ , and use of a vertical density profile over land ice. The thermal conductivity of the transition layer between the atmosphere and snow is set to that of snow and, whenever the upper snow layer exceeds 20 cm depth, it is treated as if the snow was 20 cm thick for heat calculations. In addition, sea ice albedo, sea ice thickness, and snow cover on sea ice are fixed in the standard WRF but can vary in Polar WRF since version 3.2.1. For consistency between Polar WRF versions, these last options are not used here. The distinction between the two codes has recently been blurred somewhat as some Polar WRF modifications, in particular the treatment of sea ice through the mosaic method [Bromwich *et al.*, 2009] and changes to the thermal properties of snow and ice [Hines and Bromwich, 2008] have been incorporated into the standard release from NCAR.

[10] Both models use fully compressible Euler nonhydrostatic equations for atmospheric dynamics on a horizontal Arakawa C-grid staggering and a vertical terrain-following hydrostatic-pressure coordinate. Because it is a limited-area model, Polar WRF requires information on the wind, temperature, geopotential, and relative humidity at the lateral boundaries. At the lower boundary, the surface elevation, pressure, sea ice conditions, and sea surface temperatures also have to be specified. Reflection of gravity waves from the upper boundary is problematic and is controlled by either a layer of increased diffusion, Rayleigh relaxation, or the inclusion of a layer with implicit gravity-wave damping [Chini and Leibovich, 2003]. In this study, increased diffusion and damping of the vertical velocity over an 8 km depth (approximately four levels) from the model top are applied. A more comprehensive description of WRF is available from Skamarock *et al.* [2009] and the WRF user's guide

([http://www.mmm.ucar.edu/wrf/users/docs/user\\_guide\\_V3/ARWUsers-GuideV3.pdf](http://www.mmm.ucar.edu/wrf/users/docs/user_guide_V3/ARWUsers-GuideV3.pdf)).

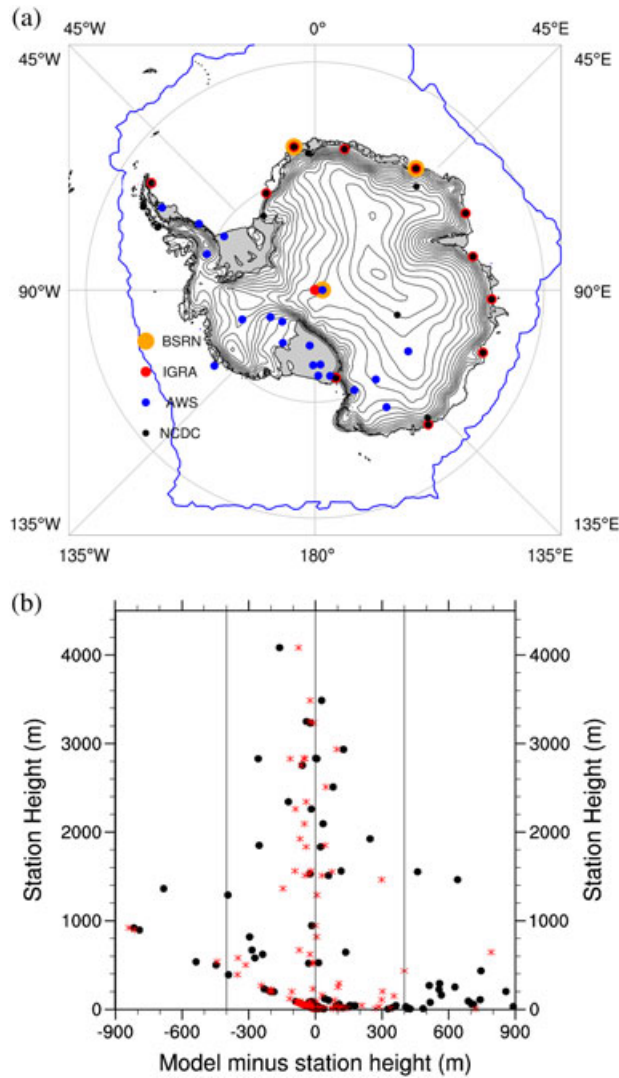
[11] Improvements in both models can be grouped under physics, dynamics, nudging, initialization, and software. With each version of WRF, not only the details of the improvements are provided but also a list of known problems. Only a summary of improvements relevant to this study is presented here. For a complete list of updates from WRF version 3.0.1 to 3.3.1, the reader should go the WRF user's website (<http://www.mmm.ucar.edu/wrf/users/>). Table 1 shows that the main change in the model versions considered here are changes to the radiation scheme; however, there are other differences in the model versions that are not immediately evident. In WRF 3.0.1, for example, an option was added that allows sea ice fraction and albedo to be updated, and some changes were made to emissivity values. In version 3.1.1, the Mellor-Yamada-Janjic (MYJ) planetary boundary layer (PBL) and MYJ surface physics were updated to the NCEP operational versions, and some minor fixes were made for polar fast Fourier transforms. The Mellor-Yamada Nakanishi Niino (MYNN) level 2.5 and level 3 PBL and MYNN surface schemes were also added. Therefore, in WRF 3.1.1 one has the option of using MYJ or MYNN for the PBL and surface physics. A fix for a cooling problem near model top that resulted from using vertically interpolated relative humidity rather than mixing ratio was made in WRF 3.2. Although all the experiments described here use the Grell-Devenyi cumulus scheme, a minor fix was added for vertical shear calculation in version 3.3.

## 2.2. Experimental Strategy

[12] In assessing the impact of model improvements, newer options for the same physics schemes are used whenever available (see Table 1 for a summary of the physics packages used). The RRTM longwave scheme [Mlawer *et al.*, 1997] is used in Polar WRF 3.0.1 and 3.1.1, whereas the RRTM for general circulation models (RRTMG) is used in Polar WRF 3.2.1 and 3.3.1. Use of the RRTMG shortwave radiation scheme is a requirement for using the new RRTMG longwave scheme. RRTMG accounts for all major gaseous absorbers at long and short wavelengths along with the radiative effects of clouds and aerosols as well as a statistical technique for representing subgrid-scale cloud variability. Hines *et al.* [2011] recently performed four simulations to assess the sensitivity of the various microphysics options in Polar WRF and found only a small sensitivity over Arctic land. Thus in Antarctica (which is mostly land ice) the impact of using the WRF Single Moment 5-class microphysics (WSM5) instead of the Morrison scheme that is used in the Arctic System Reanalysis is presumed to be small. AMPS also uses the WSM5, so the level of skill found here provides a benchmark for the comparable resolution of AMPS experimental forecasts.

[13] There are three full-year experiments using GFS-FNL analysis for 2007, but each was conducted with a different version of Polar WRF (obtained from NCAR [1999]). In addition, a total of seven sensitivity experiments described further in section 3.2 are performed for January (summer) and July (winter) 2007. They are used to assess sensitivity to lateral boundary conditions (GFS-FNL versus ERA-Interim), radiation physics, and planetary boundary layer schemes. The choice of physics experiments was based on known differences in the physical characteristics between Antarctica and





**Figure 1.** (a) Location of the observational sites superimposed on the model topography (250 m interval) and observed average sea ice extent (blue contour) depicted using the 0.01 fractional sea ice concentration contour. The blue, black, orange, and red dots denote Automatic Weather (AWS), synoptic (NCDC), radiation (BSRN), and IGRA upper air stations, respectively. Ice shelves are shaded gray. (b) Station elevation ( $y$  axis) and elevation errors ( $x$  axis) at 60 km (black dots) and at 15 km (red asterisks) model resolution; some of the stations below 0.5 km are displaced vertically in the model by nearly 1 km. Stations not properly represented in the model (height errors exceeding 400 m, indicated by vertical lines) at 60 km have been excluded from the statistical analysis.

the Arctic. Antarctic radiation balance is likely different because the clouds are composed primarily of ice and because of the high surface albedo. In addition, the thick ice sheet establishes an extremely stable boundary layer profile, which influences near-surface fluxes. Therefore, sensitivity to the PBL and radiation parameterizations were identified as being very good candidates for this study.

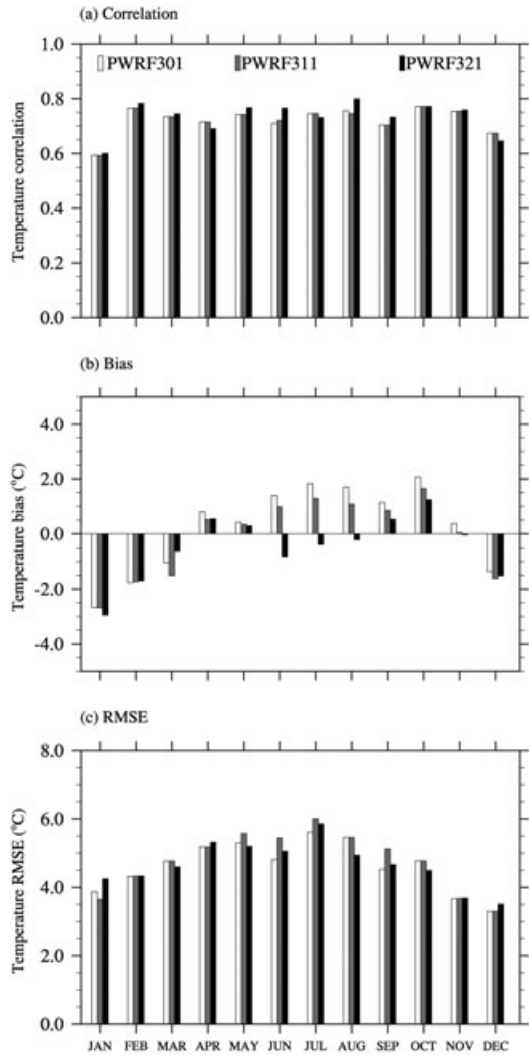
[14] The analysis and reanalysis described earlier are used to specify lateral boundary conditions on the five outermost

grid points and gradually relaxed over the next 10, resulting in a 15-point forcing frame. For each day of the year, the model is initialized with a 24 h lead time using analysis or reanalysis values (depending on the experiment) and integrated continuously for 48 h. Only forecasts from the last 24 h from each of the forecasts for the consecutive calendar days are joined to make an annual time series from which the monthly analyses are performed in section 3.0. Using only the last 24 h from each forecast allows for spin-up of the hydrological cycle and planetary boundary layer structure over the first 24 h of the 48 h forecast. *Guo et al.* [2003] used a 72 h integration period, but our preliminary analysis showed that the centers of forecast Antarctic storms drift significantly without nudging in longer-duration simulations; hence, the shorter duration is used here.

[15] Because of the amount of simulations required, a single domain comparable to that used by *Guo et al.* [2003] and centered at the South Pole with the top at 10 hPa, 39 vertical levels, and a 60 km horizontal resolution is used (Figure 1a). However, at this resolution, a number of Antarctic stations located on the steep coastal slopes and some isolated islands are not properly represented. For example, Base Orcadas (60.74°S, 44.74°W) is a land point 8 m above mean sea level, yet at this resolution the nearest grid point is located over water at sea level. The impact of the coarse resolution on the statistics is minimized in this study by using only stations whose surface elevation in the model does not differ from the actual station height by more than 0.4 km (Figure 1b) and whose weather is not strongly influenced by local features such as topography. For analysis, the forecasts are interpolated to the station sites using a bilinear interpolation while avoiding mixing land and ocean grid points because of their different diurnal characteristics. A higher resolution is achievable through nesting and could downscale the results to less than 2 km (as in AMPS Grid 5), but it is computationally expensive for our domain size, integration duration, and number of experiments. Furthermore, high forecast skill on the parent domain is critical for accuracy on any of the higher resolution nests. Nevertheless, 1 month sensitivity experiments during both summer and winter are used here to assess the impact of a fourfold increase in model resolution on forecast skill (section 3.2).

### 2.3. Lower Boundary Conditions

[16] Lower boundary conditions are specified by terrain elevation, sea ice concentrations, and sea surface temperatures (SST). Elevation data from the 200 m Radarsat Antarctic Mapping Project Digital Elevation Model (RAMP-DEM) [*Liu et al.*, 2001] are interpolated using standard WRF routines to the domain in Figure 1a. Fractional sea ice concentrations are specified from passive microwave measurements from DMSP Special Sensor Microwave Imagers (SMM/I) available from the National Snow Ice Data Center (NSIDC) [*Comiso*, 2007] and are based on the bootstrap algorithm. Although the sources of terrain and sea ice concentration remain the same for all experiments, the SST data are either specified from the archives of NCEP 0.5° real-time global (RTG) SST [*Gemmell et al.*, 2007] analyses for 2007 or the NOAA weekly 1.0° optimum interpolated observations (OI) [*Reynolds et al.*, 2007] for 1993 (prior to RTG SSTs). Both SSTs and sea ice fraction concentrations are linearly



**Figure 2.** (a–c) Domain-averaged monthly 2 m air temperature statistics; lowest correlations are found in January ( $n = 25$ ).

interpolated in time to update the lower boundary conditions every 6 h throughout the 48 h of forecast.

[17] Unlike a case in the midlatitudes where many soil temperature measurements exist, Antarctica has few observed snowpack profiles needed to specify accurately the snowpack temperature profiles. *Weller and Schwerdtfeger* [1977] measured the thermal properties of the uppermost 10 m of snow at a site on the Antarctic Plateau. The results show that the monthly temperature near the surface has a strong seasonal cycle, whereas at the 10 m depth the temperature is equal to the annual mean. Therefore, this study assumes that the stable stratification over the permanent ice cover produces ice surface temperatures that are roughly in equilibrium with the air at 2 m above the ice/snow surface. Monthly mean 2 m air temperatures are therefore used to specify the temperature at the top of the snowpack, and climatological annual mean is used to specify temperatures at the bottom of the snowpack as required at a depth of 8 m in the Noah LSM [*Chen and Dudhia*, 2001]. Temperatures for intermediate depths are linearly interpolated from these two.

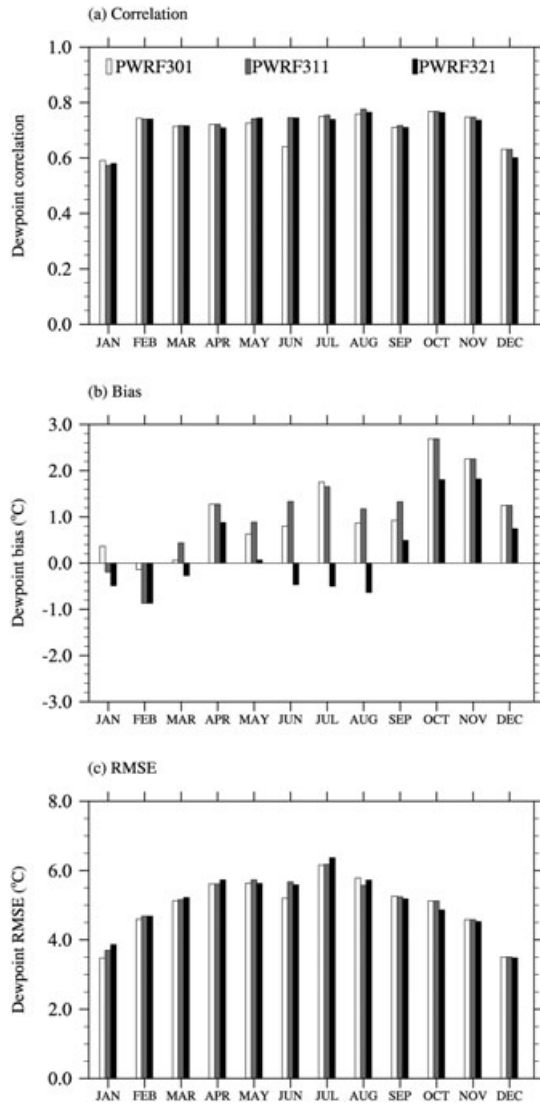
## 2.4. Observational Data

[18] The sites of observations used for the model comparison are given in Figure 1a. Synoptic station observations are retrieved from the National Climate Data Center (NCDC) archives and augmented by Automatic Weather Station (AWS) observations from the University of Wisconsin's Antarctic Meteorological Research Center (AMRC), the Italian Antarctic Research Program, and the Australian Antarctic Data Center. AWS units measure air temperature, wind speed, and wind direction at  $\sim 3$  m, whereas the other variables are measured at  $\sim 1$  m. Therefore, for AWS sites, the forecast 10 m wind speeds are adjusted using the logarithmic wind profile down to 3 m above the model surface to correspond more closely to the AWS wind measurements. Measurements of the surface downwelling shortwave and longwave radiation from the Baseline Surface Radiation Network (BSRN) [*Ohmura et al.*, 1998] of Antarctic Stations [*Dutton*, 2008; *Vitale*, 2009; *Yamanouchi*, 2010; *König-Langlo*, 2011] are used in the surface energy budget analysis. Above the surface, upper air measurements from the British Antarctic Survey and the Integrated Global Radiosonde Archive (IGRA) [*Durre et al.*, 2006, 2008] are used to evaluate the forecasts.

[19] The extreme climate and challenging environment in which Antarctic observations are made inevitably produce some unusable measurements. Therefore, very stringent quality control measures are necessary to isolate good observations. First, a visual examination individual time series for suspicious characteristics is performed for each station and variable. These include, for example, large spikes in surface pressure (e.g.,  $> 50$  hPa over a 10 min period), a zero wind speed reading with non-missing or fixed direction, etc. Because of difficulties in distinguishing AWS units reporting calm conditions from those with frozen anemometers, all wind speeds less than  $0.1 \text{ m s}^{-1}$  are treated as missing in this study. Next, the availability ( $> 50\%$ ) and representativeness ( $< \pm 3\sigma$ ) are examined. Monthly statistics are used in the quality control to accommodate Antarctic summer-only stations with extensive summer observations but with large gaps at other times. Finally, the number of stations is thinned out to prevent areas such as the Ross Ice Shelf and King George Island, where stations tend to cluster in a small area, from dominating the continental statistics. Snowfall represents a critical input to the continent's ice sheet mass balance, but accurate continent-scale observations are not available. Therefore, precipitation is not assessed in this study.

## 3. Results

[20] The problems associated with coarse resolution for Antarctic simulations were noted in section 2.4. Figure 1b shows the difference between model elevation and true station elevation as a function of elevation and demonstrates the impact of using higher model resolution. At 60 km, most of the stations have positive errors that can exceed 800 m. When the resolution is increased to 15 km (red asterisks), the errors are substantially reduced for stations below 1 km elevation. At both resolutions, some stations that are on flat ground are still located on steeply sloping model terrain. Forecast wind speed and direction errors from such misrepresentations cannot easily be corrected. For temperature and pressure, on the other hand, adjustments are made using the dry adiabatic lapse rate ( $\sim 9.8^\circ\text{C/km}$ ) and the hypsometric relationship



**Figure 3.** (a–c) Domain-averaged monthly 2 m dew point temperature statistics ( $n = 51$ ).

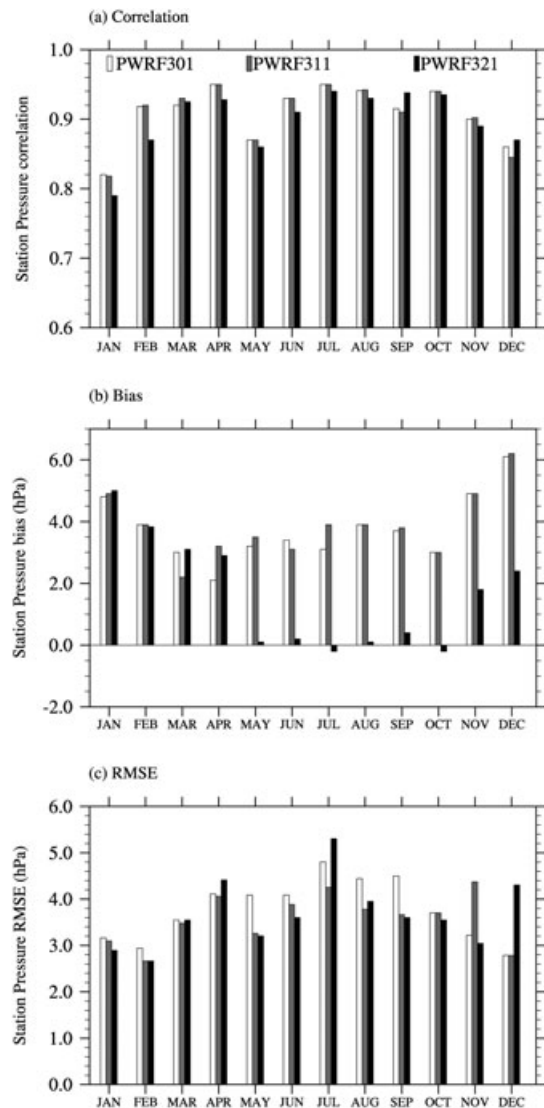
before any of the statistical analysis that follows. The mean model bias (BIAS), root mean square error (RMSE), and correlation (CORR) statistics are computed with respect to the three-hourly observations of air temperature, dew point temperature, surface pressure, wind speed and direction, and incident surface shortwave and longwave radiation.

### 3.1. Impact of Model Improvements on Forecast Skill

#### 3.1.1. Surface Air and Dew Point Temperatures

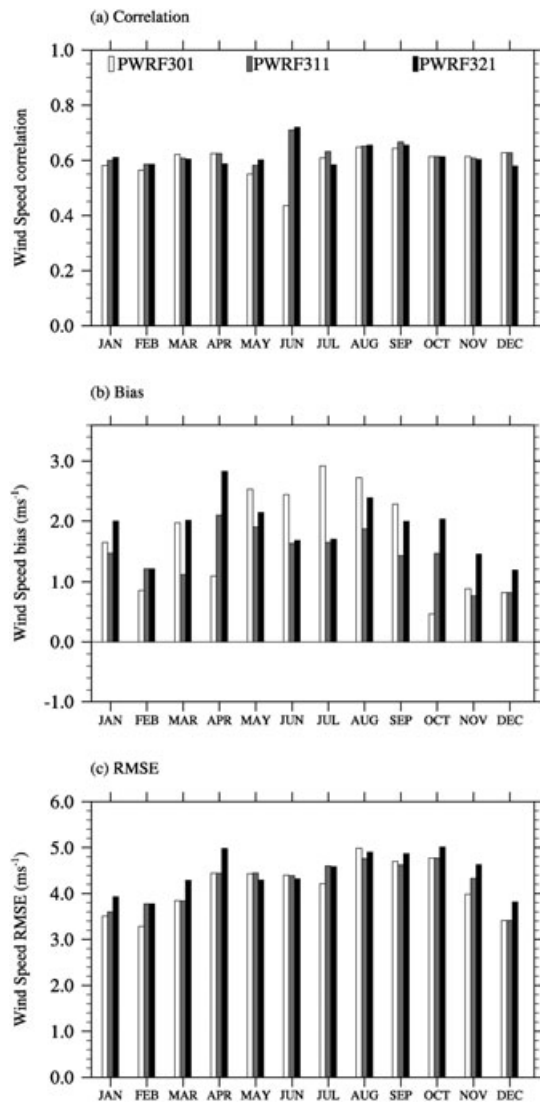
[21] The impact on forecast skill of model improvements is assessed by comparing forecasts made using GFS-FNL for a fixed period (January to December 2007) with three different versions of the model (Polar WRF 3.0.1, 3.1.1, 3.2.1; hereafter PWR301, PWR311, and PWR321, respectively). PWR311 only became available after a large number of these simulations were completed and is therefore evaluated just for January and July 2007 in the sensitivity experiments described in section 3.2. The model

physics improvements considered here are primarily in the parameterizations of shortwave and longwave radiation (Table 1). Statistics are averaged over the stations in Figure 1a for each version of the model. Figure 2 shows that the model skill in forecasting temporal variability of 2 m air temperature as measured by the temporal correlations exceeds 0.7 for most of the year. In January, a lower value of  $\sim 0.55$  indicates difficulties forecasting temporal variability at the peak of summer when the diurnal cycle is pronounced. Spatial patterns of the statistics (not shown) indicate higher correlations on the eastern plateau. Although correlations vary little between the model versions, the biases show distinct differences between the winter and the summer seasons. All versions of the model show a cold near-surface temperature bias in summer, with January peaks below  $-2^{\circ}\text{C}$ . During winter, on the other hand, a warm bias prevails, again for all three versions. The differences are greater in January ( $>2^{\circ}\text{C}$  colder) compared with July, when the warm bias remains  $<2^{\circ}\text{C}$ . Among the three versions, PWR321 generally shows



**Figure 4.** (a–c) Domain-averaged monthly station pressure statistics ( $n = 46$ ).





**Figure 5.** (a–c) Monthly mean wind speed statistics from 25 stations.

the smallest bias (both warm and cold). The RMSE differences are also larger between the seasons (up to  $1.0^{\circ}\text{C}$ ) than they are between model versions ( $<0.5^{\circ}\text{C}$ ). Stations with more variable January temperatures also tend to exhibit the largest cold bias. Large variations in the statistics for adjacent stations (expected to show similar statistics) on the Ross Ice Shelf and near the South Pole underscore the need for stringent quality control of the observations. The lowest correlations are found along the coast, another indication of the difficulty of accurately representing these stations. In summary, changes in the model versions from 3.0.1 to 3.2.1 have only a small impact on the forecasting skill of the 2 m air temperature over Antarctica.

[22] To assess forecasts of the moisture field, we use averages of monthly 2 m dew point temperatures (Figure 3). The temporal patterns are very similar to those found in the 2 m air temperature above. However, the temporal correlations are slightly lower, remaining near 0.7 throughout the year as opposed to above 0.7 for air temperature. Once again, there is little difference in the temporal correlation statistic

between model versions and hence the prediction skill of temporal variability. As with the 2 m air temperature, the models have a cold bias in dew point temperatures in January and a warm bias in July, but the cold dew point bias in January is smaller ( $<1^{\circ}\text{C}$ ) than the corresponding air temperature bias ( $>2^{\circ}\text{C}$ ). Therefore, the forecast January dew point depressions (temperature minus dew point temperature at 2 m) are smaller, indicating a forecast atmosphere closer to saturation than is observed. In spite of this, we show in section 3.1.3 that the model atmosphere has deficiencies in cloud cover that substantially impact the surface energy balance. Winter dew point temperature RMSEs are larger than those for temperature, because dew point measurements at these low temperatures are in general less reliable. The very low amounts of moisture on the continent together with difficulties in making the observation as well as in predicting their variations contribute to the larger RMSEs. The differences in seasonal correlation, bias, and RMSEs again greatly exceed those between different model versions.

### 3.1.2. Surface Pressure and Wind Speed

[23] Statistics for surface pressure are also very similar between the different model versions, and no single version consistently outperforms the others all year round. The correlations of surface pressure are generally high ( $\sim 0.9$ ), except in January and May, when they are about 0.8 (Figure 4). Although PWR301 and PWR311 show a smaller seasonal difference in the bias, a strong seasonal cycle is found in the PWR321 bias, which is also the lowest; it remains near zero in June and is negative ( $<1\text{ hPa}$ ) in July and October. Forecast surface pressures are generally higher than observed, leading to biases exceeding  $\sim 4\text{ hPa}$  in January and December for the two earlier versions. The larger January positive pressure bias is consistent with the cold temperature bias we saw earlier.

[24] Correlations for forecast monthly wind speed exceed  $\sim 0.50$  throughout the year, with no well-defined difference between winter and summer (Figure 5). Values from the different versions are again nearly equal, indicating that all three predict the temporal evolution of surface wind speed equally well. Unlike the correlations, bias between forecast and observed wind speeds shows a strong seasonal cycle ranging in magnitude from  $\sim 1\text{ m s}^{-1}$  in February to  $\sim 2.5\text{ m s}^{-1}$  in August. The peak winter wind speed bias likely is related to seasonal variation in absolute wind speeds, with higher speeds producing the larger biases. As noted above, the wind field is the most sensitive to model terrain, and inaccurate representation of complex topography cannot always be corrected easily. Therefore, in addition to the spatial averages described above, statistics from a subset of stations in relatively homogeneous subregions for January 2007 are presented in Table 2. Only PWR321 results are shown because the others are very similar. Consistent with Figure 5, average forecast wind speeds are generally higher than observed, and the forecast wind direction generally falls within the same quadrant as the observed. Also, when grouped by subregions (low elevation, ice shelf, and interior), the least wind speed biases clearly occur in the interior of the continent and the greatest occur at the low elevations, where they can exceed  $9\text{ m s}^{-1}$  both for average and for most frequent cases. These results therefore underscore the importance of choosing appropriate model resolutions, especially for domains that include the Antarctic coast, where many of the verification

**Table 2.** Surface Wind Speed Statistics from 3-Hourly Observations of a Representative Subset of Stations for January 2007<sup>a</sup>

				Avg. Speed			Avg. Direction		Most Frequent Direction		Most Frequent Speed	
		Latitude	Longitude	stddev	OBS	W321	OBS	W321	OBS	W321	OBS	W321
	<i>n</i>	(°N)	(°E)	(m/s)	(m/s)	(m/s)	(°)	(°)	(°)	(°)	(m/s)	(m/s)
Low Elevation												
Rothera	239	−67.6	−68.1	3	4	9	51	54	22	45	6	9
Syowa	238	−69.0	39.6	3	4	5	75	81	67	90	5	7
B. San Martin	220	−68.1	−67.1	7	4	9	64	73	225	90	4	15
Neumayer	244	−70.7	−8.3	4	7	7	105	96	90	90	9	10
Mirnyj	120	−66.6	93.0	3	5	14	143	104	135	90	6	15
Casey	243	−66.3	110.5	5	5	11	82	99	180	90	4	13
D-47	241	−67.4	138.7	5	12	15	154	130	157	135	13	17
Arelis	248	−76.7	163.0	3	4	12	183	208	180	225	7	14
Lola	140	−74.1	163.4	3	4	12	203	252	202	270	7	19
Rita	248	−74.7	164.0	6	7	7	272	239	270	225	10	7
Ice Shelf												
Vito	246	−78.5	177.8	2	4	6	196	184	180	180	5	6
Lorne	151	−78.3	170.0	3	4	6	222	210	202	180	7	9
Linda	247	−78.5	168.4	4	6	9	203	224	202	225	9	10
Limburt	195	−75.9	−59.3	3	4	5	196	220	180	225	5	5
Ferrell	245	−77.9	170.8	4	5	6	203	201	202	180	7	10
Interior												
Sky Blu	156	−74.8	−71.5	4	5	7	353	38	22.5	45	4	6
Nico	246	−89.0	89.7	4	4	6	308	324	337	0	5	6
Amundsen	105	−90.0	0.0	4	5	6	38	358	90	90	5	5
Henry	246	−89.0	−1.0	2	4	6	38	70	157	90	4	6
Concordia	184	−75.1	123.4	2	3	5	196	200	180	225	4	5

<sup>a</sup>The forecasts from PWR321 (W321), observed (OBS) speed, direction, and standard deviations (stddev) are computed from *n* observations in January 2007. Average wind direction is obtained using mean zonal and meridional wind speeds. The most frequent speeds and directions are based on a wind rose analysis.

stations are located. Overall, PWR321, which used the RRTMG radiation schemes, has a smaller bias in temperature, dew point, and pressure than the two earlier versions of Polar WRF.

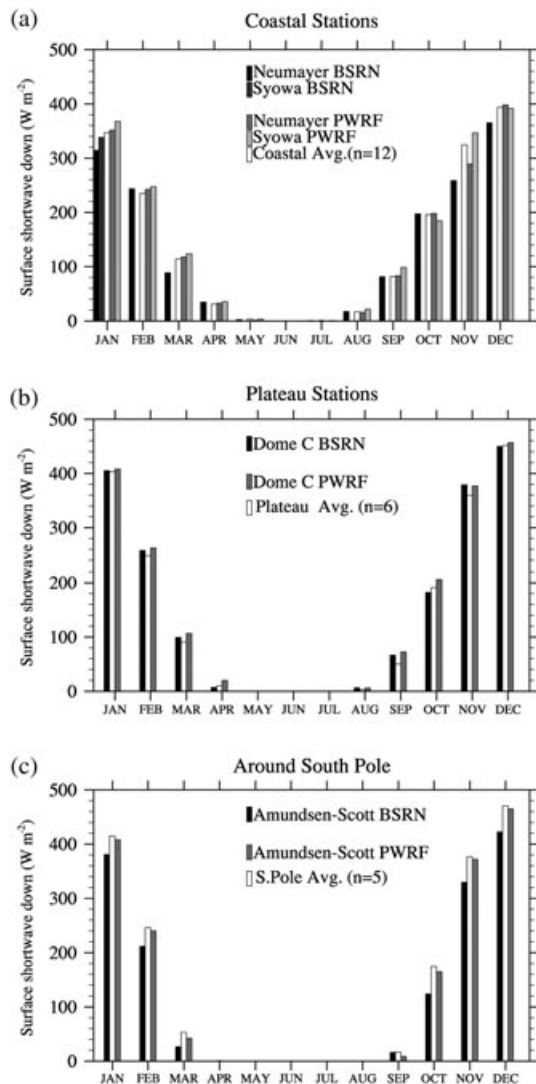
### 3.1.3. Downwelling Surface Shortwave and Longwave Radiation

[25] To determine the cause of the cold-summer and warm-winter temperature biases discussed above, we first examine the surface radiation balance using the four BSRN stations at Neumayer, Syowa, Dome C, and Amundsen-Scott. The sites range in elevation from low-altitude coastal to high-altitude locations on the interior of the East Antarctic Plateau. Figure 6 shows 2007 annual cycles of observed as well as forecast monthly incident surface shortwave (SWDOWN) radiation from PWR321 averaged over grid points that share similar geographic characteristics. In January 2007 at Neumayer (a lower latitude station), the average forecast downwelling shortwave radiation does not exceed  $360 \text{ Wm}^{-2}$ , whereas, at Amundsen-Scott (higher latitude), about  $400 \text{ Wm}^{-2}$  is simulated (Figure 6c). These forecasts of higher January insolation near the South Pole are consistent with observations showing  $\sim 380 \text{ Wm}^{-2}$  incident at the surface at Amundsen-Scott compared with only  $\sim 310 \text{ Wm}^{-2}$  at Neumayer. To show that the downwelling shortwave at Neumayer and Syowa is representative of Antarctic coastal locations, forecasts from 10 additional sites that match the locations of established coastal stations (Halley, Dumont, Casey, Mirnyj, Mawson, Rothera, Belgrano, Silvia, Mt. Siple, and McMurdo) are used. The forecast average from these sites is generally lower than the averages at Neumayer and Syowa, suggesting that there are some coastal locations where lower amounts of incident surface SWDOWN reach the surface compared with Neumayer and Syowa. However, the forecast average for the coastal

sites is still higher than the observations at Neumayer and Syowa, from November to March, when isolation plays a larger role in the surface energy budget. Therefore, on average, excessive forecasts of downwelling surface shortwave radiation at Neumayer and Syowa are characteristic of Antarctic coastal locations. A similar forecast excess is also predicted at Dome C and South Pole, but the differences are smaller. At Dome C, the forecast average from five additional locations on the East Antarctic Plateau (Vostok, Dome A, Dome Fuji, Kohonen, and AGO-3) is generally lower than observed, but the grid point closest to the station receives more downwelling shortwave radiation than observed. The average downwelling shortwave radiation around the South Pole is calculated from sites matching the locations of Henry, Nico, Panda South, AGO-4, and Amundsen-Scott. It is generally higher than both forecast and observation for the South Pole BSRN station. The diurnal cycle (not shown) indicates that the large positive SWDOWN bias occurs mostly after the local noon, especially at Neumayer. A small-amplitude forecast diurnal cycle is also found near the South Pole, because the pertinent model grid point is centered slightly off of the South Pole, but even here the bias is positive (excess) throughout the day.

[26] To assess whether the positive bias in SWDOWN is limited to 2007, a similar radiation analysis was performed for 1993. Because the full annual simulations had been completed before the release of PWR321, the results in Figure 7a are from PWR311 instead. The time series are of daytime and nighttime minimum from 1 July to 31 December followed by 1 January to 31 July, for both 1993 and 2007. This rearrangement puts the peak downwelling summer SWDOWN near the middle of the figure for clarity. Also note the lack of observed SWDOWN at Dome C in





**Figure 6.** (a–c) Mean monthly incident shortwave radiation at the surface for 2007; the average is computed for locations of similar characteristics from the forecasts. Syowa BSRN observations are available only for January.

1993. Figure 7a shows that the greater differences are found in the maximum compared with the minimum. At both Neumayer and Dome C, the forecast minimum values in shortwave are very close to the observed. The forecast maximum SWDOWN, on the other hand, is close to the observation at both Neumayer and South Pole only for larger and not smaller values of maximum SWDOWN. The high summer variability in surface downwelling shortwave at Neumayer is immediately apparent both in the observations and in the forecasts, but the observations clearly vary much more than the forecasts. Variability in the forecast maximum SWDOWN decreases more rapidly than that in observations during the fall seasons, and this can also contribute (greater averages) to the excess SWDOWN during this part of the year. Thus, larger forecasts of the smaller maximum SWDOWN values and smaller variability at higher values are responsible for the forecasts of excessive downwelling shortwave radiation.

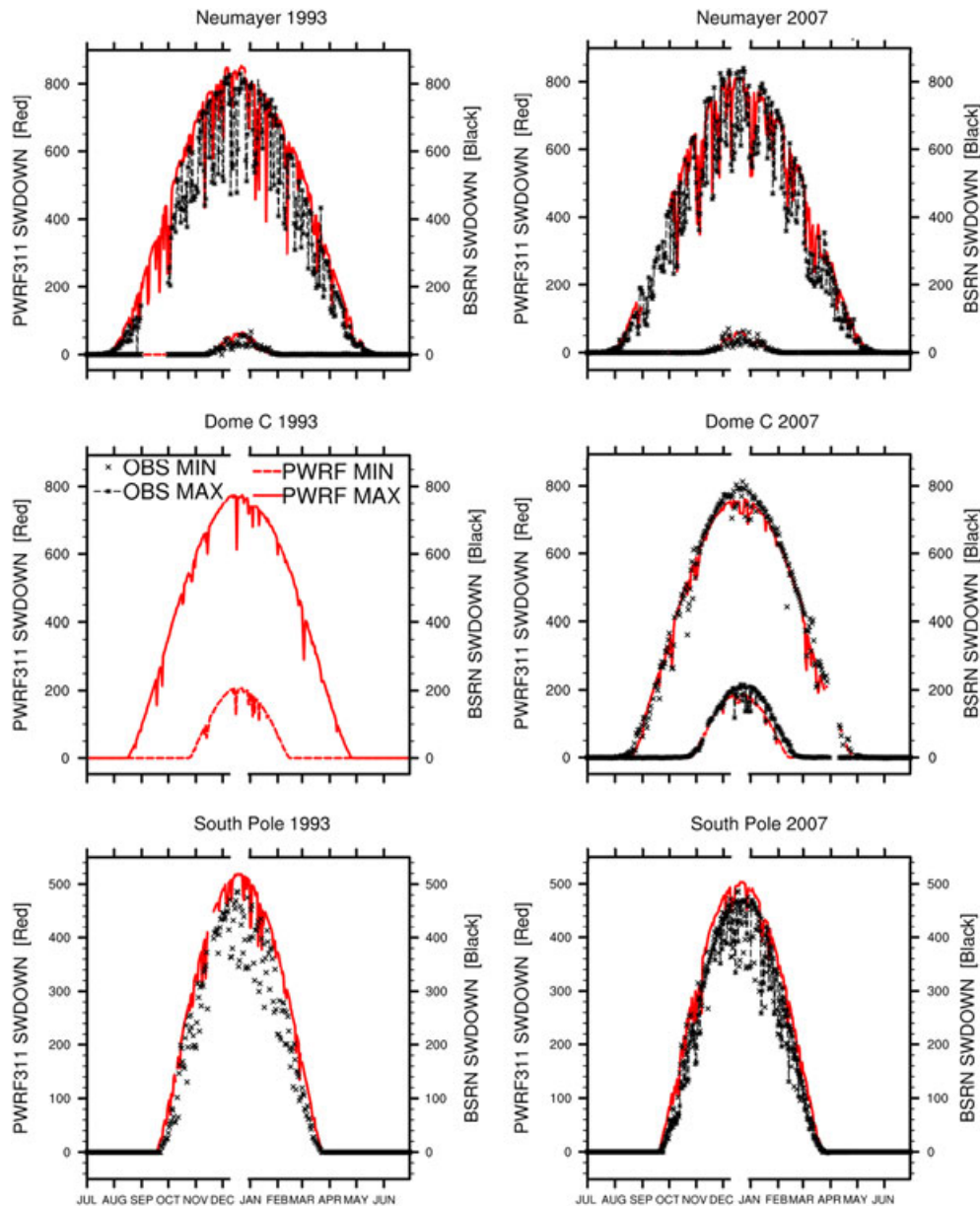
[27] To verify that cloudy skies are indeed a factor in the SWDOWN bias, forecasts of instantaneous clear-sky

downwelling shortwave radiation at South Pole were compared with those in which cloud interactions were allowed. The results (not shown) show a near-perfect match most of the time between the clear and the cloudy forecasts. Therefore, the excessive downwelling radiation results from a tendency in the model for a higher frequency of clear skies. Figure 7a also shows that the excessive SWDOWN occurs in both years and is therefore not influenced by the interannual variations in the large-scale circulation. Because the model does predict downwelling shortwave better only under clear skies, we hypothesize that inadequate cloud representation (lower amounts or smaller optical thickness) is responsible for the forecasts of excess surface incident shortwave radiation.

[28] If the model clouds are indeed deficient, a deficit in the downwelling longwave (LWD) radiation should correspond to the excess in SWDOWN. Figure 7b is a scatterplot of PWRP311 longwave down ( $y$  axis) versus the observed ( $x$  axis). In an ideal case, the values would fall exactly along the diagonal (blue line). Poor agreement between forecasts and observations is indicated by displacement from the diagonal. At Neumayer (top row), the LWD amounts cluster at both the low and the high amounts of observed LWD but show a larger spread for intermediate amounts ( $150\text{--}280\text{ Wm}^{-2}$ ) during both 1993 and 2007. Observed LWD values for 1993 are missing for South Pole and Dome C, so only 2007 values appear in the bottom row. They show clustering for lower values of LWD both at Dome C and at the South Pole. More points fall below the diagonal at the South Pole in 2007 and at Neumayer in 1993 and less distinctly in 2007. These results indicate an overall tendency for less LWD in the model than in the observations and are consistent with the positive SWDOWN bias in Figure 6. The pattern is not as striking at Dome C, which also had fewer measurements of longwave down.

[29] At South Pole and Dome C, disparities between the model and observations increase for higher LWD values (cloudy skies), further supporting a hypothesis of inadequate cloudiness. The cloud deficit was investigated using observed cloudiness from NCDC for Neumayer, Vostok, and Amundsen-Scott (Figure 8a). Observed sky cover less than 1, 1–4, 5–7, and above 7 oktas are usually classified as clear (CLR), scattered (SCT), broken (BKN), and overcast (OVC), respectively. Although making cloud observations in Antarctica during the winter darkness is difficult, the excess downwelling shortwave and deficient longwave radiation are prominent during the summer as well. At Neumayer, the frequency of overcast conditions is about 30% for all months, and the combined frequency of overcast and broken sky conditions usually exceeds 60%. This high frequency of cloudy conditions likely is due to the site's proximity to the coast and frequent passages of synoptic storms and explains the lower mean SWDOWN at Neumayer (in spite of its lower latitude; Figure 6) compared with South Pole, where the corresponding total is typically only 40%. Observations from Vostok are used instead of those from Dome C because of lack of observed cloud categories at Dome C. Clear skies occur more frequently at Vostok, also on the plateau, and we infer a similar pattern for neighboring Dome C.

[30] Performing an objective cloud comparison remains difficult because models do not forecast cloud categories and because of uncertainties in the reported cloud observations that



**Figure 7.** (a) Forecast (red) and observed (black) daytime (max) and nighttime (min) observed surface downwelling radiation ( $\text{Wm}^{-2}$ ) at (top row) Neumayer, (middle row) Dome C, and (bottom row) South Pole; only the maximum downwelling shortwave radiation is shown for South Pole because of small diurnal amplitude. Left column is for 1993 and the right column is the corresponding radiation for 2007. The gap in the  $x$ -axis indicates the discontinuity resulting from rearrangement of the time series, which places summer insolation in the middle of the plots. (b) Scatterplots of PWR311 forecast surface downwelling longwave radiation ( $\text{Wm}^{-2}$ ) for (top row) Neumayer 1993 and 2007. The bottom row is for South Pole and Dome C in 2007. The blue line marks the one-to-one line representing a perfect fit.

depend on the observer's ambient weather conditions. Differing definitions of model total cloud fraction and empirical algorithms used to infer cloud cover from satellite measurements also make it difficult to compare model cloud fraction directly against satellite measurements. Recent Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observations (CALIPSO) provide profiles of cloud liquid water path that can be more objectively compared with forecasts of the same fields from the model and therefore motivate future study.

[31] In spite of these difficulties, forecast total cloud fraction can be estimated from the atmospheric water content. The formulation used here combines model cloud liquid and ice water paths as modified for use in Antarctica by *Fogt and Bromwich* [2008]. It shows spatial patterns with higher cloud fractions at Neumayer and lower amounts toward the South Pole, in qualitative agreement with the observations described above. In Figure 8b, the derived cloud estimate is overlaid on the observed category (represented using the midrange fraction in oktas) in January 2007. Observed

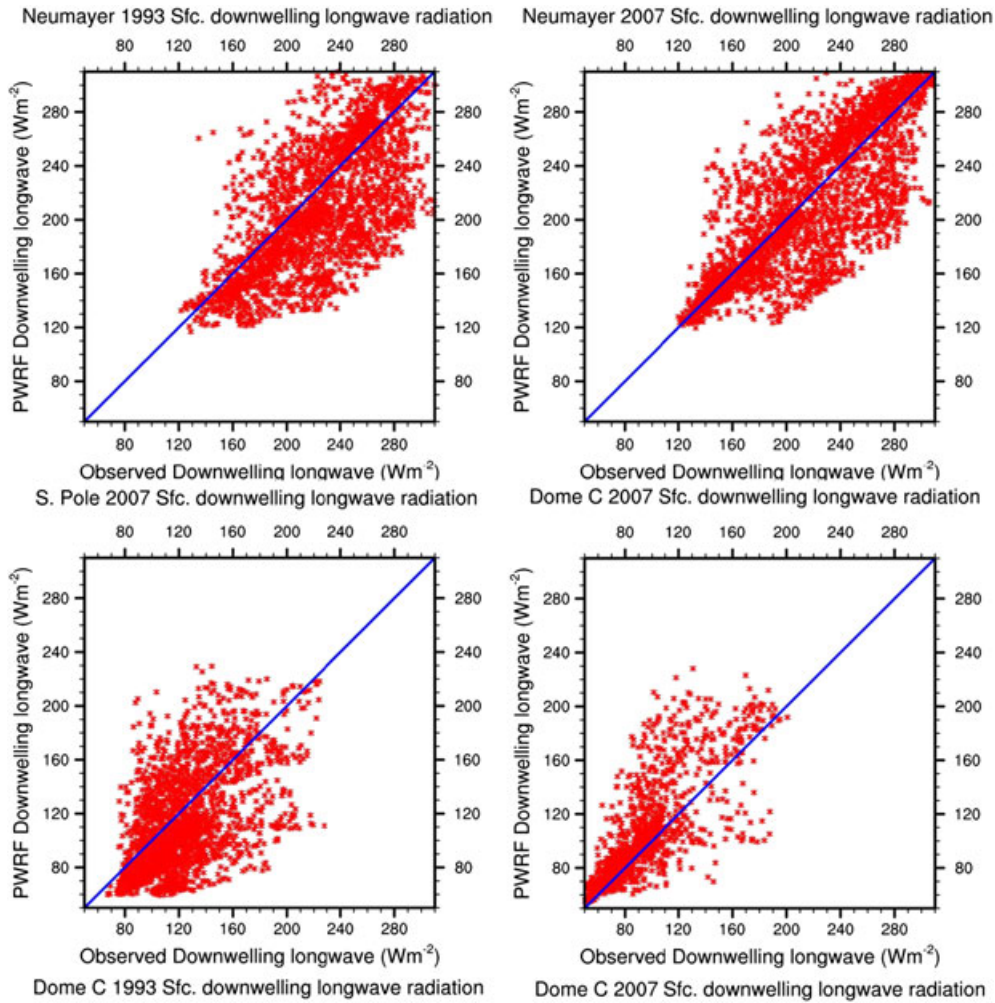


Figure 7. (continued)

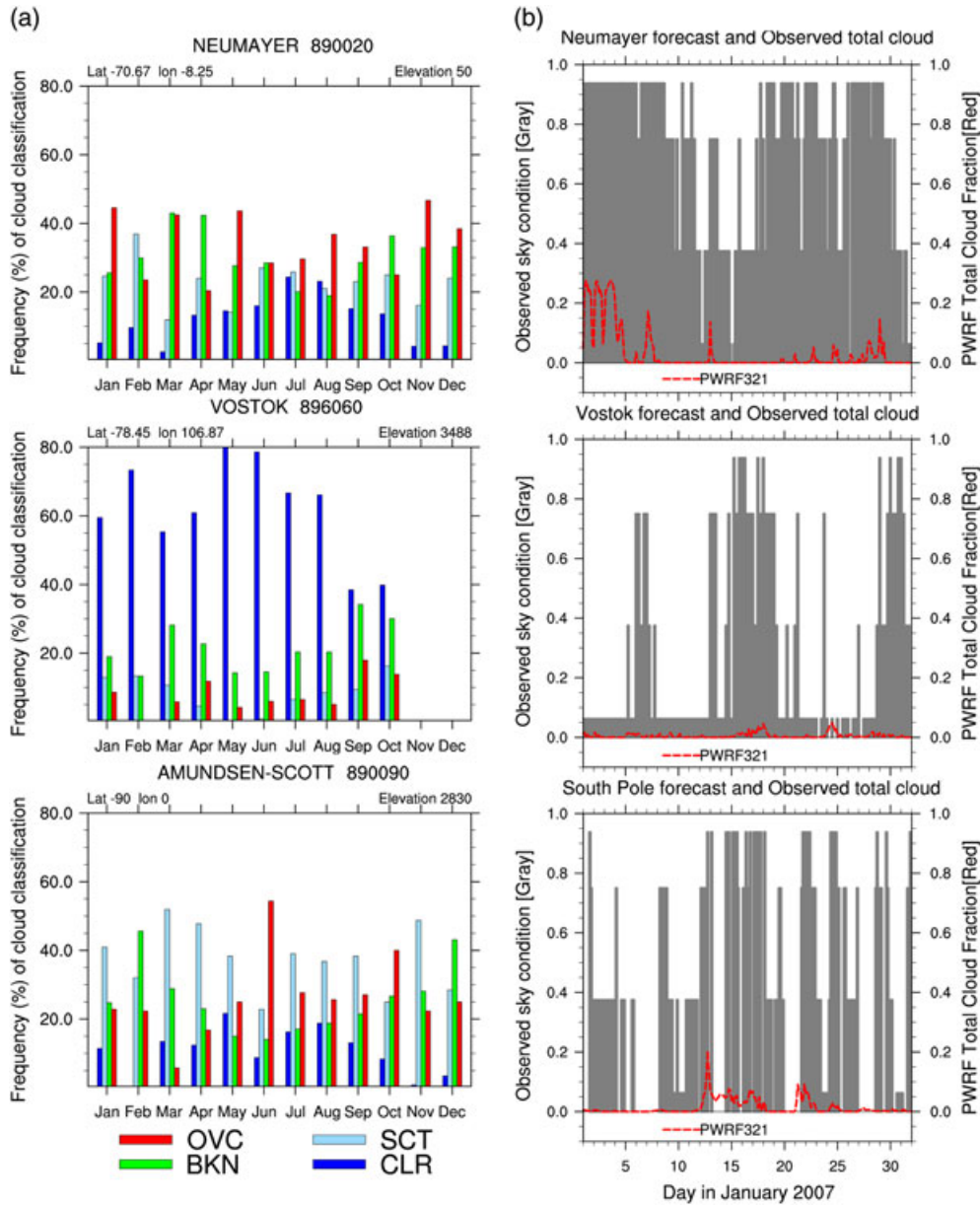
cloudier conditions (gray bars) at Neumayer relative to the other two sites are immediately obvious, and so is the fact that the computed total cloud fraction remains near zero for much of the month. Higher model total cloud fraction (red dashes) also corresponds to observed (gray bars) overcast conditions (7.5/8.0) most of the time. For intermediate observed cloudiness (3.0/8.0 and 6.0/8.0), the model cloud fraction is always close to zero. In other words, the model is able to forecast overcast events better than the broken and scattered sky conditions. Table 3 shows results from both PWR311 in 1993 and 2007 as well as PWR321 just for 2007 and therefore allows an examination of the impact of interannual differences in the surface radiation balance in addition to model improvements. The SWDOWN statistics are computed only from the sunlit part of the year (1 September to 1 March) because of the strong seasonal cycle and many zero values during the Antarctic winter. LWD statistics, on the other hand, are computed for the entire year because of the critical role LWD plays in the surface energy balance during the winter months. Except for South Pole in 1993, the temporal correlations of downwelling shortwave radiation exceed 0.9 at all three locations in both years and model versions. It is interesting that the correlation at Neumayer, where the SWDOWN variability is large, is comparable to that at both Dome C and South Pole.

Correlations for downwelling longwave are more variable and range from 0.60 to 0.80 and are lower than those of SWDOWN. This suggests that temporal variation downwelling shortwave radiation which is dominated by the diurnal cycle is better forecasted than that in LWD, which comes primarily from clouds in the atmosphere. The biases show an excess in SWDOWN at all locations but greatest at the South Pole. The excess also occurs in both years, as well as in the two model versions analyzed, but the bias in LWD is mixed, with Dome C showing more than observed surface longwave radiation for PWR311 and PWR321. The larger RMSE in both downwelling short and longwave radiation at Neumayer is not surprising given the high variability there. We conclude that the model has difficulty predicting the downward longwave radiation and hence the surface energy balance because of difficulties in correctly predicting cloud cover and its effects.

### 3.1.4. Upper Air Variables

[32] Figure 9 shows forecast and observed temperature profiles averaged from two observed profiles daily (at 0000 and 1200 UTC) available from upper air stations in the IGRA archive; therefore, the three-hourly forecasts are resampled to match the observation frequency and at the standard levels (850, 700, 500, 400, 300, 250, and 200 hPa). Table 4 gives results for all variables at 12 stations (see Figure 1a for





**Figure 8.** (a) Observed cloud frequencies from NCDC for clear (CLR), scattered (SCT), broken (BKN) and overcast (OVC) categories in 2007 at Neumayer, Vostok, and Amundsen-Scott stations. (b) Derived cloud fraction from observations (gray) and forecasts (red dashed) for January 2007. Cloudy conditions are reported at Neumayer more frequently all year, and the forecasts underestimate the total cloud cover at all three locations.

locations). Only the forecasts from PWR311 are shown, because small variations occur between the model versions in the upper air statistics and for consistency with the radiation analysis described above. In general, differences from observations in the upper air variables are smaller than at the surface. Both observed and predicted profiles clearly show the tropopause ( $\sim 300$  hPa) during summer, but the model overestimates its temperature. The model forecasts the broad characteristics of the temperature profile accurately in summer and winter, even capturing the large annual temperature range above 200 hPa ( $\sim 60$  K), which is nearly three times than that near the surface.

[33] During the summer (December, January, and February [DJF]), the temperature correlations range from 0.79 to 0.96, whereas biases remain generally cold and small except at the tropopause, where they are warmer than observed by at least  $1^\circ\text{C}$  (Table 4). In winter, the temperature correlations near the bottom of the profile are slightly higher than in summer. Unlike the surface biases discussed above, upper air temperature biases are small and remain less than  $1^\circ\text{C}$ , except at the tropopause. A cold DJF bias occurs near the bottom of the profile in all versions and is consistent with the surface temperature bias discussed earlier. The RMSE values indicate that winter temperatures are more variable, especially below

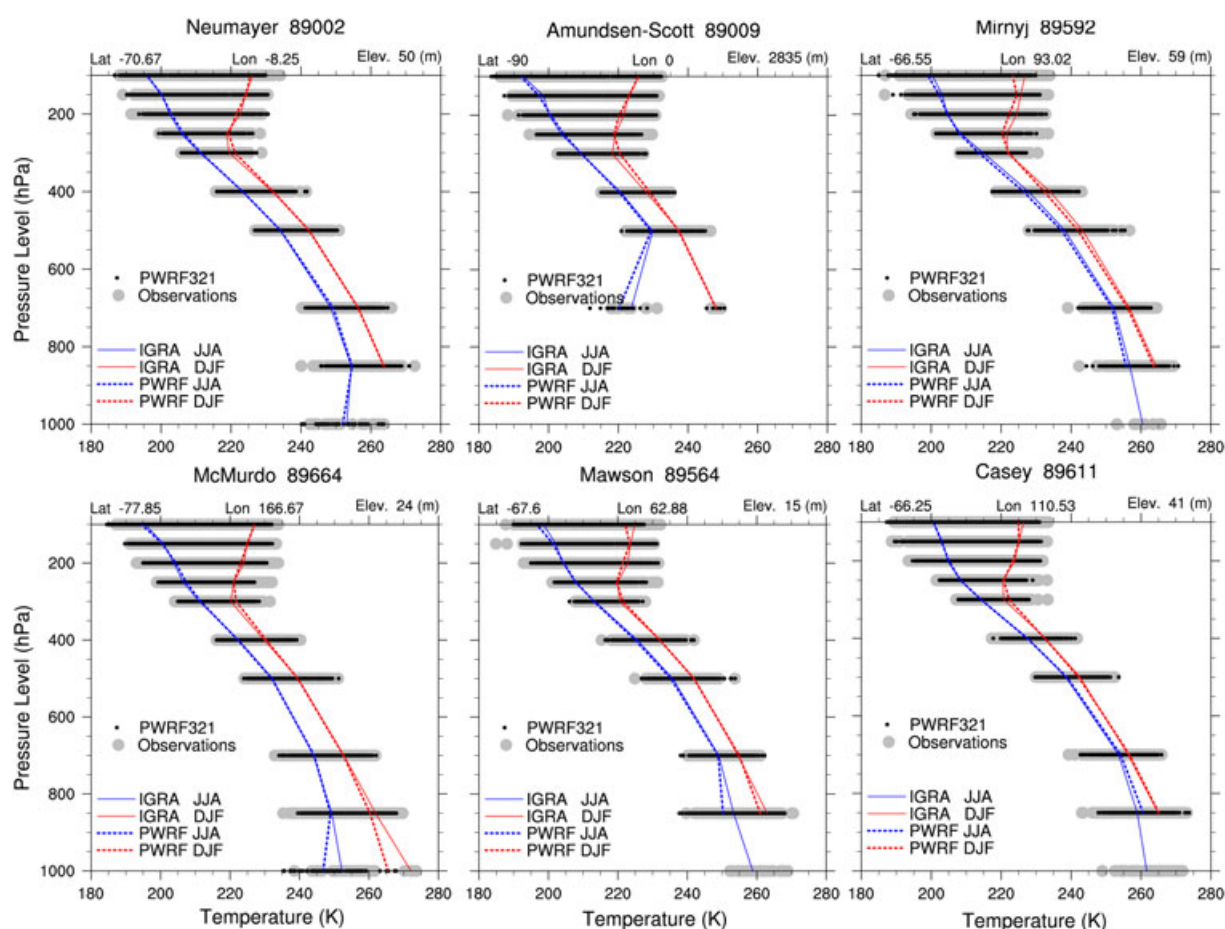
**Table 3.** Correlations, Biases, and RMSE at 3 h Intervals for Downwelling Shortwave and Longwave Radiation<sup>a</sup>

	Shortwave Down			Longwave Down		
	PWRF311		PWRF321	PWRF311		PWRF321
	1993	2007	2007	1993	2007	2007
Correlation						
Neumayer	0.94	0.95	0.94	0.65	0.77	0.76
Dome C	—	0.98	0.98	—	0.79	0.80
South Pole	0.70	0.97	0.97	—	0.63	0.60
Bias (W/m <sup>2</sup> )						
Neumayer	47.8	32.8	27.5	−18.8	−9.4	−12.9
Dome C	—	19.9	8.0	—	6.2	5.8
South Pole	53.3	66.0	54.2	—	−23.4	−23.8
RMSE (W/m <sup>2</sup> )						
Neumayer	108.0	88.7	90.9	40.6	35.8	36.3
Dome C	—	46.3	51.1	—	21.3	19.7
South Pole	73.8	58.2	46.8	—	29.2	30.4

<sup>a</sup>Statistics for downwelling shortwave radiation are for the periods 21 September to 31 December and 1 January to 21 March. Statistics for downwelling longwave radiation are for the entire year.

500 hPa, but in summer a greater variability occurs near the tropopause. Correlations between forecasts and observed geopotential heights are generally high and remain above 0.9 in both seasons. In all three versions, geopotential height biases are smaller near the bottom but increase with height to at least three times as large at 200 hPa as they are at 700 hPa. Some recent work [Cassano *et al.*, 2011; Powers *et al.*, 2010] suggests that spectral nudging near the model top or different formulation of the model layer top could reduce this error. In fact, correcting this anomaly was another motivation for using the RRTMG radiation scheme, but this was not successful.

[34] Away from the complex Antarctic topography, the upper air meridional and zonal wind speeds show higher correlations than 0.8 in both seasons above 700 hPa. Anomalous positive wind speed bias above 700 hPa is found primarily in January for the zonal component, but the zonal wind speed bias is slightly larger (absolutely) than that of the meridional component. At 850 hPa, the zonal wind is weaker than observed, whereas the meridional is stronger but the absolute bias is smaller than that found at the surface.



**Figure 9.** Observed and forecast mean vertical temperature profiles from 2007 for six Antarctic stations from IGRA; black and gray dots are forecasts and observed temperatures, respectively; blue curves are averages for the winter months (June–August), and the red curves represent summer (December–February) temperatures. Observed temperatures are continuous curves; forecasts are represented by dashed curves. Note that observations at 1000 hPa are intermittently available for Neumayer, Mirnyj, McMurdo, Mawson, and Casey and at 700 hPa for Amundsen-Scott.

**Table 4.** Domain-Averaged Upper Air Statistics for December–February Representing Summer and June–August for Winter ( $n = 12$ )

DJF				JJA			DJF			JJA		
hPa	BIAS	RMSE	CORR	BIAS	RMSE	CORR	BIAS	RMSE	CORR	BIAS	RMSE	CORR
Air temperature (K) Polar WRF3.1.1						Geopotential heights (m) Polar WRF3.1.1						
850	−0.31	1.54	0.86	0.93	2.31	0.90	1.6	15.4	0.96	7.4	17.6	0.99
700	0.05	1.54	0.88	0.34	2.01	0.88	2.3	16.0	0.92	9.4	21.6	0.94
500	−0.37	1.41	0.93	−0.55	1.50	0.93	3.3	21.1	0.97	11.3	25.1	0.98
400	−0.12	1.59	0.90	−1.01	1.60	0.93	1.2	24.4	0.97	7.7	27.1	0.98
300	1.52	2.36	0.79	−0.46	1.39	0.90	10.9	30.8	0.97	4.5	29.7	0.98
250	0.34	2.40	0.89	0.16	1.59	0.91	24.6	38.3	0.98	10.8	31.9	0.98
200	−0.63	1.68	0.96	0.68	1.84	0.95	44.4	51.8	0.98	31.4	41.4	0.99
Zonal wind speed (m/s) Polar WRF3.1.1						Meridional wind speed (m/s) Polar WRF3.1.1						
850	−0.21	3.86	0.79	−1.36	5.16	0.83	1.01	3.68	0.65	1.06	4.75	0.63
700	0.58	3.64	0.81	0.11	4.13	0.80	0.21	3.19	0.75	−1.71	3.94	0.80
500	0.77	3.70	0.84	0.61	4.29	0.90	0.13	3.68	0.87	−0.19	4.23	0.89
400	0.80	4.41	0.86	0.59	4.91	0.90	−0.07	4.76	0.87	−0.07	4.85	0.90
300	0.18	5.15	0.86	0.35	5.30	0.90	−0.09	5.70	0.88	0.07	5.20	0.91
250	−0.34	4.08	0.88	0.59	5.22	0.92	−0.05	4.44	0.91	0.21	4.47	0.92
200	0.81	2.86	0.90	−1.96	5.32	0.93	−0.89	2.89	0.93	0.29	3.65	0.93
Wind speed (m/s) Polar WRF3.1.1						Relative humidity (%) Polar WRF3.1.1						
850	0.36	4.02	0.80	1.63	5.38	0.78	1.71	17.6	0.65	−1.72	18.9	0.74
700	−0.46	3.34	0.72	0.11	4.03	0.74	1.81	19.1	0.61	−0.85	20.2	0.72
500	−0.28	4.23	0.79	−0.44	4.08	0.83	6.74	18.8	0.62	3.48	20.0	0.65
400	−0.45	5.48	0.79	−0.42	4.69	0.84	7.57	17.5	0.52	5.36	19.8	0.62
300	−2.09	4.33	0.81	−0.68	5.15	0.87						
250	−1.38	2.50	0.85	−0.95	4.77	0.89						
200	0.05	1.98	0.92	−0.46	5.26	0.89						

Statistics for the total wind speed (Table 4) show similar characteristics, higher winter correlations and marked height dependence in both seasons. Relative humidity statistics show the lowest skill of all the variables. Only values for levels below 400 hPa are shown, because the measurements are less reliable, with many missing observations higher up. The highest summer and winter correlations occur below 700 hPa and are 0.65 and 0.74, respectively. Biases are positive in summer and mixed in winter and are generally less than 5%.

### 3.2. Sensitivity to Physics Schemes, Interannual Variability, and Resolution

[35] This section discusses experiments in which the model sensitivity to interannual variations, radiation physics, PBL schemes, uncertainty in driving data, and horizontal model resolution are evaluated. Only January and July are analyzed because they showed the largest seasonal contrast. Improvement in the observation network in Antarctica is evident from occurrence of more than twice as many useable observations in 2007 as in 1993. The uncertainty for each month is calculated by the ratio of the standard deviation to the square root of the number of stations used from that month. An evaluation of the latest version of Polar WRF (3.3.1) is included in Table 5 (column 7). This version was released after many of the experiments described above had been completed.

[36] The first two columns in Table 5 show statistics from the two different years (1993 and 2007) for the same configuration of PWR311 driven by the same analysis (ERA-Interim). The summer temperature correlation, RMSE, and bias are all higher in 2007 than in 1993, but the aforementioned cold bias is exhibited in both years. A similar pattern is seen in the surface pressure and wind speed statistics but not in dew point, for which the correlations, RMSE, and bias are higher in 1993 than in 2007. A cold summer and stronger wind speed bias are common to both Polar WRF forecasts.

During July, the surface temperature bias is cold in these runs driven by ERA-Interim, in sharp contrast to the warm bias found in the earlier experiments driven by GFS-FNL. This finding indicates that the model is very sensitive to the data used to provide the lateral boundary and initial conditions.

[37] The next three columns in Table 5 assess the sensitivity of PWR321 to radiation and PBL physics, so the GFS-FNL data are used in all three experiments. Temperature correlations are comparable, and the biases are negative in summer and positive in July, in agreement with the full-year experiments that also used GFS-FNL described in section 3.1. The correlations are also higher during July for all three. The run with the community atmospheric model (CAM) radiation has the least winter 2 m temperature bias. Dew point temperatures show the least summer skill in all three experiments, but the winter skill is slightly better than the corresponding summer values. Wind speed statistics show that all three experiments have stronger than observed wind speeds in both months and that, for each month, the individual statistics from the experiments are comparable. Replacing the MYJ PBL scheme in PWR321 with the MYNN PBL [Nakanishi and Niino, 2004] scheme (columns 3 and 5 in Table 5) increases the summer cold bias but does not substantially change the other statistics. The final two columns in Table 5 compare the newest version of Polar WRF (3.3.1) with the preceding version (3.2.1). In these two experiments, the driving data and physics combinations are exactly the same.

[38] The most striking feature when the results from experiments with ERA-Interim are compared with those with GFS-FNL is the drastic reduction in both the summer and the winter temperature biases to near zero. A possible explanation is that the GFS-FNL analysis initializes Polar WRF to a state that is too cold for the model to recover from within a 48 h period. A similar improvement is found in the wind speed statistics, in which a positive bias still exists but is noticeably smaller. To compare objectively all the experiments from 2 years



**Table 5.** January and July Correlations (CORR), RMSE, and Bias From the Sensitivity Experiments ( $n$  = Number of Stations)<sup>a</sup>

Model/Year	PWRF311 1993	PWRF311 2007	PWRF321 2007	PWRF321 2007	PWRF321 2007	PWRF321 2007	PWRF331 2007
Physics	RTM/MYJ	RTM/MYJ	RRTMG/MYJ	CAM/MYJ	RRTMG/MYN	RRTMG/MYN	RRTMG/MYN
Analysis	ERA-Interim	ERA-Interim	GFS-FNL	GFS-FNL	GFS-FNL	ERA-Interim	ERA-Interim
<i>January</i>							
2 m Temp. ( $n = 23, 65$ )							
CORR	0.57 ( $\pm 0.03$ )	0.63 ( $\pm 0.03$ )	0.61 ( $\pm 0.02$ )	0.61 ( $\pm 0.02$ )	0.59 ( $\pm 0.02$ )	0.61 ( $\pm 0.02$ )	0.62 ( $\pm 0.02$ )
RMSE (K)	3.59 ( $\pm 0.47$ )	4.11 ( $\pm 0.19$ )	3.55 ( $\pm 0.17$ )	3.85 ( $\pm 0.19$ )	4.70 ( $\pm 0.25$ )	3.53 ( $\pm 0.17$ )	3.41 ( $\pm 0.17$ )
BIAS (K)	-1.83 ( $\pm 0.55$ )	-3.05 ( $\pm 0.26$ )	-1.28 ( $\pm 0.26$ )	-2.12 ( $\pm 0.26$ )	-3.32 ( $\pm 0.30$ )	0.09 ( $\pm 0.30$ )	-0.06 ( $\pm 0.29$ )
Stn. Pressure ( $n = 17, 68$ )							
CORR	0.92 ( $\pm 0.02$ )	0.97 ( $\pm 0.01$ )	0.96 ( $\pm 0.01$ )	0.93 ( $\pm 0.01$ )	0.96 ( $\pm 0.01$ )	0.97 ( $\pm 0.01$ )	0.95 ( $\pm 0.01$ )
RMSE (hPa)	3.60 ( $\pm 0.60$ )	4.23 ( $\pm 0.26$ )	4.68 ( $\pm 0.28$ )	5.45 ( $\pm 0.28$ )	4.64 ( $\pm 0.28$ )	4.25 ( $\pm 0.24$ )	4.98 ( $\pm 0.25$ )
BIAS (hPa)	0.14 ( $\pm 0.66$ )	0.57 ( $\pm 0.45$ )	0.01 ( $\pm 0.48$ )	-0.08 ( $\pm 0.47$ )	-0.06 ( $\pm 0.47$ )	0.73 ( $\pm 0.41$ )	1.33 ( $\pm 0.44$ )
2 m Dew Point ( $n = 17, 30$ )							
CORR	0.55 ( $\pm 0.05$ )	0.51 ( $\pm 0.04$ )	0.46 ( $\pm 0.04$ )	0.43 ( $\pm 0.04$ )	0.51 ( $\pm 0.04$ )	0.44 ( $\pm 0.04$ )	0.44 ( $\pm 0.04$ )
RMSE (K)	5.45 ( $\pm 1.11$ )	3.57 ( $\pm 0.24$ )	4.73 ( $\pm 0.32$ )	4.45 ( $\pm 0.29$ )	4.03 ( $\pm 0.25$ )	5.52 ( $\pm 0.40$ )	5.41 ( $\pm 0.39$ )
BIAS (K)	2.03 ( $\pm 0.82$ )	1.30 ( $\pm 0.39$ )	2.93 ( $\pm 0.40$ )	2.16 ( $\pm 0.41$ )	1.15 ( $\pm 0.43$ )	4.03 ( $\pm 0.46$ )	3.92 ( $\pm 0.46$ )
Wind Speed ( $n = 7, 18$ )							
CORR	0.24 ( $\pm 0.04$ )	0.60 ( $\pm 0.04$ )	0.55 ( $\pm 0.05$ )	0.54 ( $\pm 0.04$ )	0.55 ( $\pm 0.05$ )	0.59 ( $\pm 0.05$ )	0.61 ( $\pm 0.04$ )
RMSE (m/s)	4.37 ( $\pm 0.80$ )	3.53 ( $\pm 0.44$ )	3.67 ( $\pm 0.41$ )	3.70 ( $\pm 0.43$ )	3.68 ( $\pm 0.44$ )	3.43 ( $\pm 0.41$ )	3.35 ( $\pm 0.38$ )
BIAS (m/s)	0.69 ( $\pm 1.18$ )	1.60 ( $\pm 0.38$ )	1.40 ( $\pm 0.39$ )	1.52 ( $\pm 0.38$ )	1.64 ( $\pm 0.39$ )	1.07 ( $\pm 0.39$ )	0.86 ( $\pm 0.39$ )
Rank score	68	43	54	59	52	42	33
<i>July</i>							
2 m Temp. ( $n = 26, 67$ )							
CORR	0.77 ( $\pm 0.03$ )	0.78 ( $\pm 0.01$ )	0.77 ( $\pm 0.01$ )	0.77 ( $\pm 0.02$ )	0.77 ( $\pm 0.01$ )	0.78 ( $\pm 0.02$ )	0.75 ( $\pm 0.02$ )
RMSE (K)	6.23 ( $\pm 1.49$ )	5.89 ( $\pm 0.24$ )	5.65 ( $\pm 0.24$ )	5.56 ( $\pm 0.18$ )	5.70 ( $\pm 0.24$ )	5.65 ( $\pm 0.24$ )	5.72 ( $\pm 0.27$ )
BIAS (K)	-1.37 ( $\pm 0.80$ )	-1.01 ( $\pm 0.44$ )	1.69 ( $\pm 0.37$ )	0.11 ( $\pm 0.41$ )	1.87 ( $\pm 0.37$ )	0.89 ( $\pm 0.38$ )	-0.17 ( $\pm 0.40$ )
Stn. Pressure ( $n = 18, 65$ )							
CORR	0.95 ( $\pm 0.0$ )	0.97 ( $\pm 0.01$ )	0.97 ( $\pm 0.01$ )	0.94 ( $\pm 0.01$ )	0.97 ( $\pm 0.03$ )	0.97 ( $\pm 0.01$ )	0.95 ( $\pm 0.01$ )
RMSE (hPa)	3.06 ( $\pm 0.84$ )	4.08 ( $\pm 0.23$ )	4.51 ( $\pm 0.23$ )	5.26 ( $\pm 0.24$ )	4.47 ( $\pm 0.23$ )	4.10 ( $\pm 0.21$ )	4.85 ( $\pm 0.22$ )
BIAS (hPa)	0.37 ( $\pm 0.82$ )	0.91 ( $\pm 0.41$ )	0.43 ( $\pm 0.44$ )	0.36 ( $\pm 0.42$ )	0.36 ( $\pm 0.43$ )	1.08 ( $\pm 0.37$ )	1.67 ( $\pm 0.40$ )
2 m Dew Point ( $n = 17, 24$ )							
CORR	0.52 ( $\pm 0.05$ )	0.79 ( $\pm 0.02$ )	0.76 ( $\pm 0.02$ )	0.75 ( $\pm 0.03$ )	0.76 ( $\pm 0.02$ )	0.78 ( $\pm 0.02$ )	0.76 ( $\pm 0.03$ )
RMSE (K)	6.21 ( $\pm 1.25$ )	6.04 ( $\pm 0.33$ )	6.63 ( $\pm 0.41$ )	5.76 ( $\pm 0.34$ )	6.67 ( $\pm 0.41$ )	6.33 ( $\pm 0.34$ )	6.13 ( $\pm 0.33$ )
BIAS (K)	2.24 ( $\pm 0.91$ )	1.25 ( $\pm 0.79$ )	3.77 ( $\pm 0.62$ )	2.13 ( $\pm 0.59$ )	3.92 ( $\pm 0.60$ )	2.81 ( $\pm 0.66$ )	1.61 ( $\pm 0.70$ )
Wind Speed ( $n = 7, 14$ )							
CORR	0.32 ( $\pm 0.06$ )	0.58 ( $\pm 0.05$ )	0.55 ( $\pm 0.05$ )	0.56 ( $\pm 0.03$ )	0.55 ( $\pm 0.04$ )	0.54 ( $\pm 0.05$ )	0.53 ( $\pm 0.09$ )
RMSE (m/s)	0.80 ( $\pm 0.60$ )	4.65 ( $\pm 0.52$ )	4.62 ( $\pm 0.47$ )	5.09 ( $\pm 0.51$ )	4.62 ( $\pm 0.47$ )	4.47 ( $\pm 0.49$ )	4.49 ( $\pm 0.53$ )
BIAS (m/s)	3.90 ( $\pm 0.80$ )	1.88 ( $\pm 0.48$ )	2.11 ( $\pm 0.41$ )	2.82 ( $\pm 0.44$ )	2.12 ( $\pm 0.39$ )	1.76 ( $\pm 0.39$ )	1.36 ( $\pm 0.48$ )
Rank score	40	36	45	45	75	38	47

<sup>a</sup>Physics combinations used and whether the experiment was driven by GFS-FNL or ERA-Interim lateral boundary conditions (LBC) are also indicated. Enclosed in parentheses are the estimated ranges of uncertainty in the sample statistics. The uncertainty for each month is calculated by the ratio of the standard deviation to the square root of the number of stations used from that month. The rank score is the sum of the ranks (high correlation low rank; high bias and RMSE, low ranks) for each experiment from each variable. High overall skill corresponds to low rank score.

with different numbers of observations and disparate physics combinations and driving data, rank scores are computed. Correlations for each variable are ranked across the experiments from the largest to the smallest and corresponding RMSE and bias from the smallest to the largest; individual ranks are then summed to define the rank scores shown at the bottom of Table 5. PWRF331 and PWRF311 have the best overall performance in January and July, respectively, but PWRF321 does well in both months.

[39] Section 3 has shown vertical displacement of the location of some stations in the model as a result of coarse resolution. An additional sensitivity experiment was conducted to examine how the model skill would be impacted by resolution higher than 60 km. The experiment in the column under PWRF331 in Table 5 was repeated at the 15 km resolution used by AMPS for the Antarctic continent. Table 6 compares statistics for low-elevation stations (below 1 km), where elevation errors were largest. Except for the fourfold increase in horizontal resolution, the two sensitivity experiments are otherwise identical. For 2 m air temperature, the January correlation decreases from 0.66 to 0.64, and the bias decreases from -1.24 to -0.62 K. During July, on the

other hand, correlations are in general lower at 15 km. This possibly is due to strong synoptic influence in July, which may not benefit from the higher resolution. Statistics for the surface pressure show the least sensitivity to the increased model resolution in both months. The forecast statistics for dew point temperature show higher skill in the 15 km run in both January and July, having higher correlations and smaller bias and RMSE. The largest sensitivity to model resolution is found in the forecast for January wind speeds. The correlation increases, whereas both the RMSE and the bias decrease. In fact, the wind speed bias is reduced by 50% or more in both January and July. The results show that, although high resolution is beneficial, even higher resolution may be needed to capture fully the complex coastal environment.

[40] Scatterplots shown in Figure 10 compare experiments with PWRF321 (columns 4 and 5 of Table 5) that use the RRTMG and CAM radiation and MYJ and MYNN PBL physics, respectively. Figure 10 allows for an evaluation of model sensitivity in areas without observed data in the Southern Ocean and over Antarctica and also depicts the range of model-to-model difference over the entire domain. The top row shows the impact of analysis (ERA-Interim

**Table 6.** Average Statistics for Low-Elevation (<1 km) Stations at 60 and 15 km for January and July 2007<sup>a</sup>

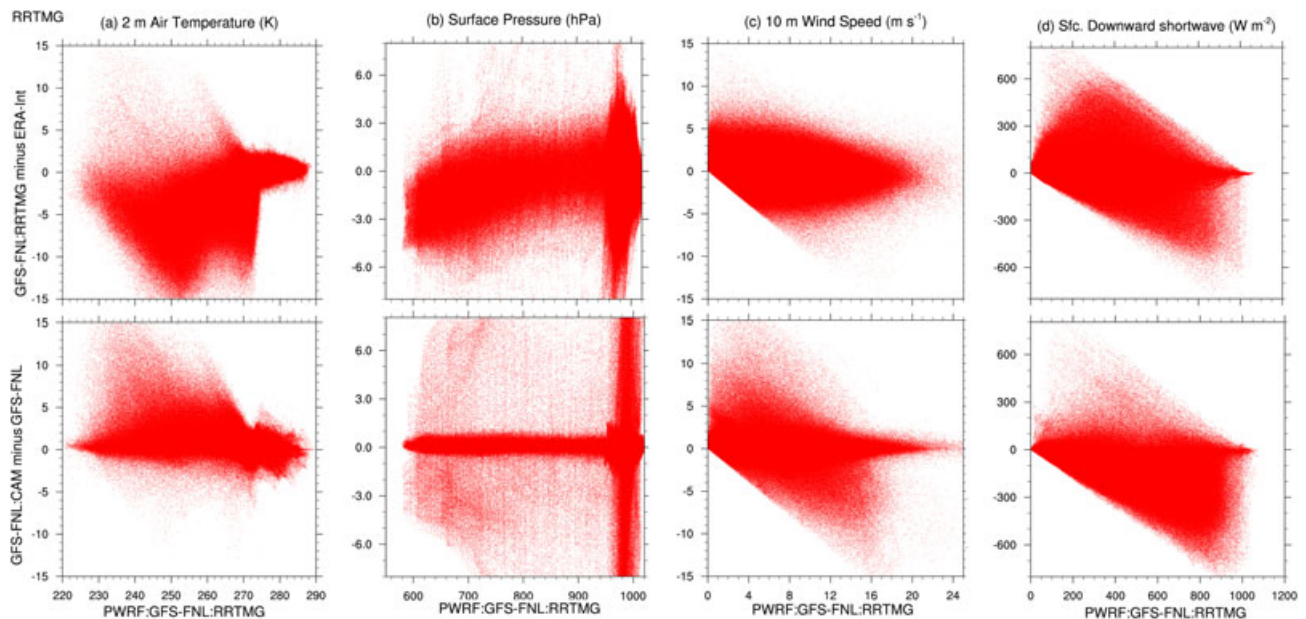
PWR331	January 2007		July 2007	
	60 km	15 km	60 km	15 km
2 m Temp. ( $n=35, 30$ )				
CORR	0.66	0.64	0.81	0.74
RMSE (K)	3.29	3.22	5.53	6.03
BIAS (K)	-1.24	-0.62	-0.59	-0.22
Stn. Pressure ( $n=33, 33$ )				
CORR	0.97	0.97	0.97	0.97
RMSE (hPa)	2.38	2.22	3.84	3.31
BIAS (hPa)	-0.53	-0.62	0.17	0.18
2 m Dew Point ( $n=20, 21$ )				
CORR	0.56	0.58	0.80	0.84
RMSE (K)	3.09	3.03	6.54	5.56
BIAS (K)	0.53	0.59	-1.04	0.38
Wind Speed ( $n=28, 24$ )				
CORR	0.65	0.67	0.58	0.58
RMSE (m/s)	3.04	2.81	5.24	4.53
BIAS (m/s)	0.95	0.20	2.52	1.30

<sup>a</sup>Except for the fourfold increase in horizontal resolution, the two sensitivity experiments are identical. The two numbers enclosed in parenthesis are the number of stations used for January and July. At each resolution, exactly the same stations are used, but not all have observations in both months.

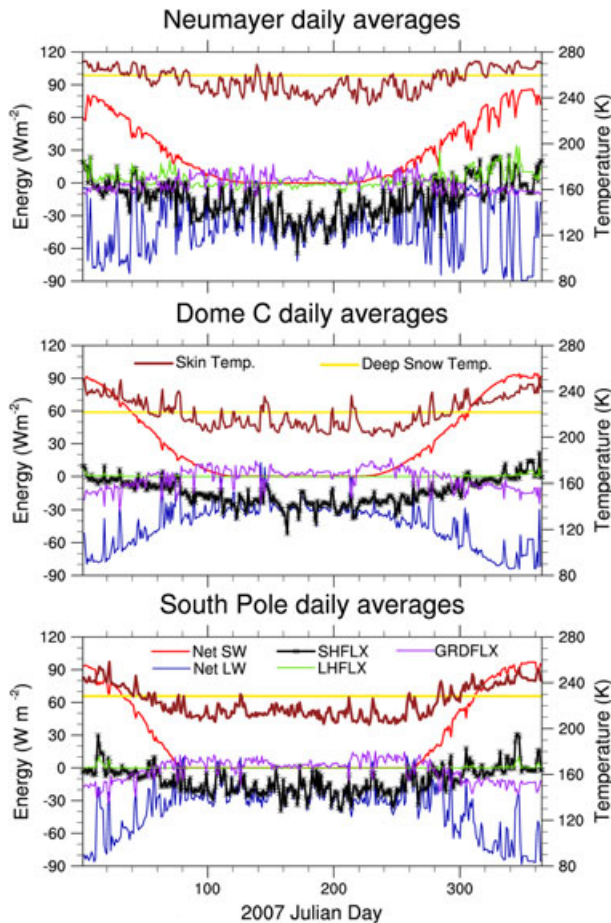
versus GFS-FNL), and the bottom row depicts model sensitivity to the radiation physics (CAM/MYJ versus RRTMG/MYNN), because these two PBL schemes produce very similar results (Table 5). The differences ( $y$  axis) are calculated at each model grid point for January 2007 and then plotted against the value ( $x$  axis) for that grid point from PWR321 forecasts of 2 m temperature, surface pressure, wind speed, and downwelling shortwave radiation (Figures 10a–10d, respectively). More locations are colder in the GFS-FNL run than in the ERA-Interim reanalysis run for temperatures

below 270 K. More locations in the CAM/MYJ experiment are warmer (reduced the colder summer bias) than in the RRTMG experiment for temperatures in the 230–270 K range. In both cases, the temperature differences become smaller for temperatures above 274 K. The surface pressure scatterplot (Figure 10b) shows larger surface differences near 1000 hPa (over the ocean primarily) and smaller values corresponding to lower pressures (over elevated Antarctica). Below 850 hPa, forecast surface pressures from the GFS-FNL run are generally lower than those from the ERA-Interim experiment. This could result from differences in the locations of synoptic systems between the two analyses in these experiments. For CAM/MYJ versus RRTMG/MYNN experiments, which both use GFS-FNL, the surface pressure differences are almost symmetrical about zero (bottom row in Figure 10b). Consistent with statistics shown above, the wind speeds are generally stronger in the GFS-FNL experiments relative to the ERA-Interim experiment (Figure 10c). Figure 10 also shows that, between the analysis- and the reanalysis-driven experiments, positive differences are more frequent for lower to intermediate wind speeds but negative for speeds exceeding  $10 \text{ m s}^{-1}$ . Weaker winds (less than  $5 \text{ m s}^{-1}$ ) are over predicted especially in the GFS-FNL run. Switching the radiation physics does not cause as much spread in the wind speed as does switching the driving analysis.

[41] Figure 10d shows that the GFS-FNL run forecasts more surface downwelling shortwave radiation than the ERA-Interim for values less than  $400 \text{ W m}^{-2}$ . Switching the radiation physics to CAM reduces this bias but has only a modest impact for shortwave radiation exceeding  $600 \text{ W m}^{-2}$ . Figure 10 also shows that, above  $800 \text{ W m}^{-2}$ , differences between the experiments can approach  $600 \text{ W m}^{-2}$ . Thus the changes in incident surface shortwave radiation resulting



**Figure 10.** January 2007 scatterplots of Polar WRF 3.2.1 differences with (top row) GFS-FNL minus ERA-Interim runs and (bottom row) GFS-FNL with CAM minus GFS-FNL with RRTMG. The scatterplots show the range of differences found at Polar WRF grid points between two sensitivity experiments for (a) 2 m temperature, (b) surface pressure, (c) wind speed, and (d) downwelling surface shortwave radiation.



**Figure 11.** PWR321 annual cycle of 2007 daily averaged net shortwave (SW, red) and longwave (LW, blue) radiation; surface sensible (SHFLX, black/asterisk), latent (LHFLX, green), and ground (GRDFLX, magenta) heat fluxes at the three BSRN sites. The sign convention for SHFLX and LHFLX is positive upward into the atmosphere. GRDFLX is negative if downward from the surface. The scale at right is for the skin (maroon) and deep snowpack (yellow) temperatures.

from a change in physics have a smaller (smaller differences) impact on the forecast SWDOWN compared with changing the driving analysis (top row in Figure 10d). From this analysis we conclude that Polar WRF is more sensitive where the 2 m air temperatures are below 270 K (over the ice sheet) and where surface pressures are above 950 hPa (coastal regions and Southern Ocean) and that the model is more sensitive to the driving analysis than a change in radiation scheme. This suggests that, although comparisons with observations discussed above yield average model statistics that are comparable between different model versions, the model is sensitive to the conditions over the Southern Ocean, where observations are scarce and a poor representation of the synoptic systems could be detrimental to the forecast over the ice sheet.

#### 4. Discussion

[42] The main findings can be summarized as follows. The forecast skill of Polar WRF improved only slightly between versions 3.0.1 and 3.3.1. Forecast 2 m temperatures exhibit a cold summer and warm winter bias, and the wind speeds

are stronger than observed. A deficit in the downwelling longwave resulting from inadequate cloud representation is inferred to be responsible for the summer cold bias. Qualitative support for this hypothesis is found when observed and estimated cloudiness are compared. The excess in downwelling shortwave radiation plays a secondary role, because  $\sim 80\%$  is reflected back and lost to space with the high surface albedo and clear model skies in Antarctica. To demonstrate that deficit in downwelling longwave radiation explains most of the summer surface temperature bias, we use the surface energy budget equation. Assuming a balance in which the net shortwave radiation, sensible, latent, and snowpack heat flux all remain unchanged, the temperature change that results from increasing the downward longwave radiation based on the deficits from Table 3 is calculated using the Stefan Boltzmann equation. The results show an increase in summer surface temperatures of 1.4 and 2.9 K at Neumayer and South Pole, respectively. An increase of 2 K represents more than 60% of the cold bias (Table 5a) in the GFS-FNL runs.

[43] Recall that the summer cold bias is almost eliminated when ERA-Interim reanalysis is used in the sensitivity experiments described in section 3.2. This indicates that the analysis used is also an important factor in the cold bias and most likely is communicated through the input moisture field. ERA-Interim sensitivity experiments with a smaller cold bias show higher moisture flux into the atmosphere and hence more atmospheric water vapor, a well-known greenhouse gas across the continent. Many radiation schemes currently in use rely on relative humidity, cloud water, or cloud ice to diagnose cloud-radiation interaction, yet we have found relative humidity to be one of the fields predicted with the least skill.

[44] Surface energy fluxes at locations corresponding to the BSRN sites are shown in Figure 11, as are the skin temperatures and the temperatures at 8 m depth. The 8 m temperature remains constant throughout the year and is equal to the annual mean for 2007. However, the skin temperature shows both diurnal and seasonal variations. During summer, the skin temperature is generally warmer than at 8 m depth, resulting in a reversal of the snowpack temperature gradient between the seasons. As a result of our treatment of the snowpack temperature profiles (section 2.3), the heat fluxes within the snowpack are small and play only a secondary role in the temperature biases discussed above. Without this treatment, however, an extremely sharp temperature gradient develops in the top layer of the snowpack and enhances the downward heat flux, producing a fairly rapid temperature increase at depths below 2 m within the 48 h integration period.

[45] The net shortwave radiation fluxes are positive during summer, whereas the net longwave radiation is negative (energy loss) all year round. Figure 11 shows a net radiative heating (net shortwave + net longwave radiation) at Neumayer. The ground heat flux exceeds  $10 \text{ Wm}^{-2}$  at Dome C and South Pole but not at Neumayer. Both the latent and the snowpack heat fluxes are small in July, so the balance during this time is between the fluxes of downward sensible heat and net longwave radiation loss from the surface. The warm winter bias discussed above must therefore come primarily from excessive downward sensible heat fluxes caused by too much mixing in the stable Antarctic boundary layer associated with the positive wind speed bias. Stronger than



observed wind speeds appear to be a common problem for stable boundary layers with all PBL schemes used here as well as with the Yonsei PBL (C. Mass, personal communication, 2011). The sensible heat flux along the coast can be substantially larger than on the plateau as a result of mechanically induced mixing supported by the anomalously strong forecast winds. We hypothesize that the model winter fluxes of sensible heat are larger than they should be, but observed estimates in Antarctica vary too much to be useful in evaluating these simulations. Town and Walden [2009] compiled estimates of sensible and latent heat fluxes at or near the South Pole. Observed July sensible heat fluxes (negative downwards) range from about  $-5$  to  $-25 \text{ W m}^{-2}$ . The model predicted sensible heat flux directed toward the surface from the sensitivity runs range from  $-20.8$  to  $-29.8 \text{ W m}^{-2}$  and is therefore close to the highest observational estimate.

[46] The results described here have a winter bias (warm) similar to that of the shorter integrations of *Tastula and Vihma* [2011], who used a different analysis (ERA-40 reanalysis) and year (1998) in their comparison. Using 28 sites in an evaluation of Polar MM5 forced with ECMWF TOGA, Guo *et al.* [2003] found an average cold bias of  $\sim 2^\circ\text{C}$  throughout the year, together with positive pressure, wind speed, and mixing ratio biases, and proposed yearlong series of short integrations to evaluate the model further. Such integrations have been completed in this study and show comparable statistics. Thus, we can conclude that the statistics presented here are representative of the skill of Polar WRF and that the temperature and wind speed biases identified are a recurrent problem in Antarctic forecasts. These must be addressed because they affect the simulated snowdrift during the cold months and ice melt during the melt season and therefore the Polar WRF simulated surface mass balance. It is important for these factors to be simulated accurately for reliable prediction of long-term changes in ice sheet surface mass balance.

## 5. Conclusions

[47] We have examined three factors that can influence the skill of Polar WRF in Antarctic forecasts. The results show that recent versions of Polar WRF have skill that is comparable to that of a recent Arctic simulation [Wilson *et al.*, 2011] and marginally better than that of an earlier Antarctic simulation with Polar MM5 [Guo *et al.*, 2003]. The model skill varies seasonally, reflecting the different roles played by local and synoptic scale systems at different times of the year. The model shows a cold surface temperature bias in the summer months but a warm bias during winter months, when insolation plays a negligible role in heating the surface. The model over predicts the summer incident surface shortwave and under predicts the longwave down at the surface. Forecast skill is better away from the complex topography. The lower skill near the surface is due partially to the 60 km resolution used here, but the summer cold bias is driven primarily by deficits in downwelling longwave radiation. No nudging was done in this study, although it has been suggested [Cassano *et al.*, 2011] that spectral nudging near the model top can improve the performance of Polar WRF. The statistics presented here therefore provide a pure forecast benchmark of the skill of Polar WRF in Antarctic simulations. Comparison of the statistics from different years, physics

schemes, and analyses shows that the model skill is affected more by the analysis used than by the parameterization physics or year-to-year differences in the atmospheric circulation. Thus improving the network of verification observations and quality of Antarctic analyses must remain a top priority for further development of numerical modeling in Antarctica.

[48] **Acknowledgments.** This work was funded by NASA grant NNX08AN57G and AMPS grants NSF ANT-1135171, NSF OPP-0838967, and NSF ANT-1049089. The authors appreciate the support of the Antarctic Meteorological Research Center for the providing the AWS data (Matthew Lazzara and Elena Willmot, NSF grant ANT-0838834). Aaron Wilson provided highly appreciated and valuable discussions from his insight from working with Polar WRF in the Arctic.

## References

- Bracegirdle, T. J., and G. J. Marshall (2012), The reliability of Antarctic tropospheric pressure and temperature in the latest global reanalyses, *J. Clim.*, **25**, 7138–7146, doi:10.1175/JCLI-D-11-00685.1.
- Bromwich, D. H., and Z. Liu (1996), An observational study of the katabatic wind confluence zone near Siple Coast, West Antarctica, *Mon. Weather Rev.*, **124**, 462–477.
- Bromwich, D. H., J. J. Cassano, T. Klein, G. Heinemann, K. M. Hines, K. Steffen, and J. E. Box (2001), Mesoscale modeling of katabatic winds over Greenland with the Polar MM5, *Mon. Weather Rev.*, **129**, 2290–2309.
- Bromwich, D. H., K. M. Hines, and L.-S. Bai (2009), Development and testing of Polar WRF: 2. Arctic Ocean, *J. Geophys. Res.*, **114**, D08122, doi:10.1029/2008JD010300.
- Bromwich, D. H., D. F. Steinhoff, I. Simmonds, K. Keay, and R. L. Fogt (2011a), Climatological aspects of cyclogenesis near Adelie Land, Antarctica, *Tellus*, **63A**(5), 921–938, doi:10.1111/j.1600-0870.2011.00537.x.
- Bromwich, D. H., J. P. Nicolas, and A. J. Monaghan (2011b), An assessment of precipitation changes over Antarctica and the Southern Ocean since 1989 in contemporary global reanalyses, *J. Clim.*, **24**, 4189–4209, doi:10.1175/2011JCLI4074.1.
- Cassano, J. J., and T. R. Parish (2000), An analysis of the nonhydrostatic dynamics in numerically simulated Antarctic katabatic flows, *J. Atmos. Sci.*, **57**, 891–898, doi:10.1175/1520-0469(2000)057<0891:AAOTND>2.0.CO;2.
- Cassano, J. J., M. E. Higgins, and M. W. Seefeldt (2011), Performance of the Weather Research and Forecasting (WRF) Model for month-long Pan-Arctic simulations, *Mon. Weather Rev.*, **139**, 3469–3488, doi:10.1175/MWR-D-10-05065.1.
- Chen, F., and J. Dudhia (2001), Coupling an advanced land-surface/hydrology model with the Penn State/NCAR MM5 modeling system. Part I: Model implementation and sensitivity, *Mon. Weather Rev.*, **129**, 569–585, doi:10.1175/1520-0493(2001)129<0569:CAALSH>2.0.CO;2.
- Chini, G. P., and S. Leibovich (2003), An analysis of the Klemp and Durran radiation boundary condition as applied to dissipative internal waves, *J. Phys. Oceanogr.*, **33**, 2394–2407, doi:10.1175/1520-0485(2003)033<2394:AAOTKA>2.0.CO;2.
- Comiso, J. (1999; updated 2007), Bootstrap sea ice concentrations from NIMBUS-7 SMMR and DMSP SSM/I [2007]. Boulder, Colorado, National Snow and Ice Data Center. Digital media.
- Dee, D. P., et al. (2011), The ERA-Interim reanalysis: configuration and performance of the data assimilation system, *Q. J. R. Meteorol. Soc.*, **137**(656), 553–597, doi:10.1002/qj.828.
- Durre, I., R. S. Vose, and D. B. Wuertz (2006), Overview of the Integrated Global Radiosonde Archive, *J. Clim.*, **19**, 53–68, doi:10.1175/JCLI3594.1.
- Durre, I., R. S. Vose, and D. B. Wuertz (2008), Robust automated quality assurance of radiosonde temperatures, *J. Appl. Meteorol. Climatol.*, **47**, 2081–2095, doi:10.1175/2008JAMC1809.1.
- Dutton, E. G. (2008), Basic and other measurements of radiation at station South Pole (2007-05), *Climate Monitoring & Diagnostics Laboratory*, Boulder, doi:10.1594/PANGAEA.706267.
- EPICA Community Members (2004), Eight glacial cycles from an Antarctic ice core, *Nature*, **429**, 623–628, doi:10.1038/nature02599.
- Fogt, R. L., and D. H. Bromwich (2008), Atmospheric moisture and cloud cover characteristics forecast by AMPS, *Weather Forecast.*, **23**, 914–930, doi:10.1175/2008 WAF2006100.1.
- 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 260, 39 pp. NOAA, Silver Spring, Md.
- Guo, Z., D. H. Bromwich, and J. J. Cassano (2003), Evaluation of Polar MM5 simulations of Antarctic atmospheric circulation, *Mon. Wea. Rev.*, **131**, 384–411, doi:10.1175/1520-0493(2003)131<0384:EOPMSO>2.0.CO;2.

- Hannon, J. (2012), Op Deep Freeze 2011–2012 season concluded, Americas North shore journal, <http://northshorejournal.org/op-deep-freeze-2011-2012-season-concluded>.
- Hines, K. M., and D. H. Bromwich (2008), Development and testing of Polar WRF. Part I. Greenland Ice Sheet meteorology, *Mon. Weather Rev.*, **136**, 1971–1989, doi:10.1175/2007MWR2112.1.
- Hines, K. M., D. H. Bromwich, L.-S. Bai, M. Barlage, and A. G. Slater (2011), Development and testing of Polar WRF. Part III. Arctic land, *J. Clim.*, **24**, 26–48, doi:10.1175/2010JCLI3460.1.
- IAATO (2010), Antarctic Tourism fact sheet <http://iaato.org/tourism-statistics>.
- König-Langlo, G. (2011), Basic and other measurements of radiation at Neumayer Station (2007–03), *Alfred Wegener Institute for Polar and Marine Research, Bremerhaven*, doi:10.1594/PANGAEA.759355.
- Laine, V. (2008), Antarctic ice sheet and sea ice regional albedo and temperature change, 1981–2000, from AVHRR Polar Pathfinder data, *Remote Sensing Environ.*, **112**, 646–667.
- Liu, H., K. Jezek, B. Li, and Z. Zhao (2001), Radarsat Antarctic Mapping Project digital elevation model version 2. Boulder, CO: National Snow and Ice Data Center. Digital media.
- Mlawer, E. J., S. J. Taubman, P. D. Brown, M. J. Iacono, and S. A. Clough (1997), RRTM, a validated correlated-k model for the longwave, *J. Geophys. Res.*, **102**, 16,663–16,682.
- Monaghan, A. J., D. H. Bromwich, H. Wei, A. M. Cayette, J. G. Powers, Y.-H. Kuo, and M. Lazzara (2003), Performance of weather forecast models in the rescue of Dr. Ronald Shemanski from South Pole in April 2001, *Weather Forecast.*, **18**, 142–160.
- Monaghan, A. J., D. H. Bromwich, J. G. Powers, and K. W. Manning (2005), The climate of the McMurdo, Antarctica, region as represented by one year of forecasts from the Antarctic Mesoscale Prediction System, *J. Clim.*, **18**, 1174–1189.
- Nakanishi, M., and H. Niino (2004), An improved Mellor–Yamada level-3 model with condensation physics: Its design and verification, *Boundary Layer Meteorol.*, **112**, 1–31.
- National Research Council (2011), Future science opportunities in Antarctica and the Southern Ocean, <http://www.scribd.com/doc/64382324/Future-Science-Opportunities-in-Antarctica-and-the-Southern-Ocean#archive>.
- NCAR (1999), Data set ds083.2: NCEP FNL operational model global tropospheric analyses, <http://dss.ucar.edu/datasets/ds083.2>, Boulder, Colo.
- Nicolas, J. P., and D. H. Bromwich (2011), Precipitation changes in high southern latitudes from global reanalyses: A cautionary tale, *Surveys in Geophysics*, **32**, 475–494, doi:10.1007/s10712-011-9114-6.
- Nigro, M. A., J. J. Cassano, M. A. Lazzara, and L. M. Keller (2012), Case study of a barrier wind corner jet off the coast of the Prince Olav Mountains, Antarctica, *Mon. Weather Rev.*, **140**, 2044–2063.
- NSF (2012), Antarctic medical evacuation, Press release 12-149.
- Ohmura, A., and Coauthors (1998), Baseline Surface Radiation Network (BSRN/WRMC), a new precision radiometry for climate research, *Bull. Am. Meteorol. Soc.*, **79**, 2115–2136.
- Parish, T. R., and D. H. Bromwich (2007), Re-examination of the near-surface air flow over the Antarctic continent and implications on atmospheric circulations at high southern latitudes, *Mon. Weather Rev.*, **135**, 1961–1973.
- Powers, J. G. (2007), Numerical prediction of an Antarctic Severe Wind Event with the Weather Research and Forecasting (WRF) model, *Mon. Weather Rev.*, **135**, 3134–3157.
- Powers, J. G., S. M. Cavallo, and K. W. Manning (2010), Improving upper-level performance in AMPS: Longwave radiation, 5th Antarctic Meteorological Observation, Modeling, and Forecasting Workshop, 12–14 July 2010, BPRC Columbus, OH.
- Powers, J., K. W. Manning, D. H. Bromwich, J. J. Cassano, and A. M. Cayette (2012), A decade of Antarctic science support through AMPS, *Bull. Am. Meteorol. Soc.*, **93**, 1699–1712.
- Reynolds, R. W., T. M. Smith, C. Liu, D. B. Chelton, K. S. Casey, and M. G. Schlax (2007), Daily high-resolution blended analyses for sea surface temperature, *J. Clim.*, **20**, 5473–5496, doi:10.1175/2007JCLI1824.1.
- Rignot, E., I. Velicogna, M. R. van den Broeke, A. Monaghan, and J. Lenaerts (2011), Acceleration of the contribution of the Greenland and Antarctic ice sheets to sea level rise, *J. Geophys. Res.*, **38**, doi:10.1029/2011GL046583.
- Seefeldt, M. W., and J. J. Cassano (2012), A description of the Ross Ice Shelf air stream (RAS) through the use of self-organizing maps (SOMs), *J. Geophys. Res.*, **117**, D09112, doi:10.1029/2011JD016857.
- Skamarock, W. C., J. B. Klemp, J. Dudhia, D. O. Gill, D. M. Barker, X.-Y. Huang, W. Wang, and J. G. Powers (2009), A description of the Advanced Research WRF Version 3, *NCAR Technical Note NCAR/TN-475 + STR*, 113 pp.
- Steinhoff, D. F. (2011), Dynamics and variability of Foehn winds in the McMurdo Dry Valleys, Antarctica, PhD dissertation, Atmospheric Sciences Program, The Ohio State University, Columbus, Oh, 271 pp.
- Steinhoff, D. F., D. H. Bromwich, and A. J. Monaghan (2012), Dynamics of the foehn mechanism in the McMurdo Dry Valleys of Antarctica from Polar WRF, *Quart. J. Roy. Meteor. Soc.*, doi: 10.1002/qj.2038.
- Tastula, E.-M., and T. Vihma (2011), WRF Model experiments on the Antarctic atmosphere in winter, *Mon. Weather Rev.*, **139**, 1279–1291, doi:10.1175/2010MWR3478.1.
- Town, M. S., and V. P. Walden (2009), Surface energy budget over the South Pole and turbulent heat fluxes as a function of an empirical bulk Richardson number, *J. Geophys. Res.*, **114**, D22107, doi:10.1029/2009JD011888.
- Velicogna, I. (2009), Increasing rates of ice mass loss from the Greenland and Antarctic ice sheets revealed by GRACE, *Geophys. Res. Lett.*, **36**, L19503, doi:10.1029/2009GL040222.
- Vitale, V. (2009), Basic measurements of radiation at Concordia station (2006–02), *Institute of Atmospheric Sciences and Climate of the Italian National Research Council, Bologna*, doi:10.1594/PANGAEA.711756.
- Weller, G., and P. Schwerdtfeger (1977), Thermal properties and heat transfer processes of low-temperature snow. Antarctic meteorological studies at plateau station, Antarctica, *Antarctic Res. Ser.*, **25**, 27–39.
- Wilson, A. B., D. H. Bromwich, and K. M. Hines (2011), Evaluation of Polar WRF forecasts on the Arctic System Reanalysis domain. I. Surface and upper air analysis, *J. Geophys. Res.*, **116**, D11112, doi:10.1029/2010JD015013.
- Wilson, A. B., D. H. Bromwich, and K. M. Hines (2012), Evaluation of Polar WRF forecasts on the Arctic System Reanalysis domain. 2. Atmospheric hydrologic cycle, *J. Geophys. Res.*, **117**, D04107, doi:10.1029/2011JD016765.
- Yamanouchi, T. (2010), Basic and other measurements of radiation at station Syowa (2007–04), *National Institute of Polar Research, Tokyo*, doi:10.1594/PANGAEA.740932.