Front. Earth Sci.
https://doi.org/10.1007/s11707-022-0971-8

RESEARCH ARTICLE

Polar WRF V4.1.1 simulation and evaluation for the Antarctic and Southern Ocean

Jianjun XUE[1,2,4], Ziniu XIAO (✉)[1], David H. BROMWICH (✉)[3], Lesheng BAI[3]

1 State Key Laboratory of Numerical Modeling for Atmospheric Sciences and Geophysical Fluid Dynamics, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029, China
2 University of Chinese Academy of Sciences, Beijing 100049, China
3 Polar Meteorology Group, Byrd Polar and Climate Research Center, The Ohio State University, Columbus OH 43210, USA
4 China Meteorological Administration Training Centre, Beijing 100081, China

Abstract A recent version of the Polar Weather Research and Forecasting model (Polar WRF) has been upgraded to the version 4.X era with an improved NoahMP Land Surface Model (LSM). To assess the model performance over the Antarctic and Southern Ocean, downscaling simulations with different LSM (NoahMP, Noah), WRF versions (Polar WRF 4.1.1 and earlier version 4.0.3, WRF 4.1.1), and driving data (ERA-Interim, ERA5) are examined with two simulation modes: the short-term that consists of a series of 48 h segments initialized daily at 0000 UTC with the first 24 h selected for model spin-up, whereas the long-term component used to evaluate long-term prediction consists of a series of 38−41 day segments initialized using the first 10 days for spin-up of the hydrological cycle and planetary boundary layer structure. Simulations using short-term mode driven by ERA-Interim with NoahMP and Noah are selected for benchmark experiments. The results show that Polar WRF 4.1.1 has good skills over the Antarctic and Southern Ocean and better performance than earlier simulations. The reduced downward shortwave radiation bias released with WRF 4.1.1 performed well with PWRF411. Although NoahMP and Noah led to very similar conclusions, NoahMP is slightly better than Noah, particularly for the 2 m temperature and surface radiation because the minimum albedo is set at 0.8 over the ice sheet. Moreover, a suitable nudging setting plays an important role in long-term forecasts, such as reducing the surface temperature diurnal cycle near the coast. The characteristics investigated in this study provide a benchmark to improve the model and guidance for further application of Polar WRF in the Antarctic.

Received September 27, 2021; accepted January 17, 2022

E-mail: xiaozn@lasg.iap.ac.cn (Ziniu XIAO)
bromwich.1@osu.edu (David H. BROMWICH)

Keywords Polar WRF, downscaling simulation, performance evaluation, the Antarctic and Southern Ocean

1 Introduction

Based on decades of research, it has become increasingly apparent that the Antarctic and Southern Ocean play a significant role in the past, present, and future of the global climate system. This region not only responds to global change but also is the origin of important processes that control or modulate global water, heat, energy and chemistry budgets (Mayewski et al., 2009; Kennicutt et al., 2014, 2015, 2016) Concerns about these changes, including but not limited to, ice loss from the West Antarctic Ice Sheet (WAIS), the contribution of Antarctica to global sea-level and changes in atmospheric/ocean circulation, have resulted in numerous studies that demonstrated the critical factors (e.g., Monaghan and Bromwich, 2008; Bromwich et al., 2012, 2013a, 2014; Chen et al., 2014; DeConto and Pollard, 2016; Scambos et al., 2017; Scott et al., 2019; Zou et al., 2019). Therefore, it is necessary to develop accurate numerical tools to help understand the physical/dynamic/thermal processes and evaluate/predict their future impacts. The representation of polar key processes in the atmosphere, ocean, sea ice, and land surface; promotion of model formulation; high resolution; and development of coupled model systems; also need to be refined (Jung et al., 2015; Scambos et al., 2017). Fortunately, many advances have been achieved in improving numerical simulations by weather and climate models during the last decade. These developments include better model processes (parameterization schemes), higher horizontal and vertical resolution for the model, more observational

data, powerful supercomputers, high-speed communication networks, etc. Thence, accurate initial conditions and numerical models with better forecasting skill have become available (e.g., Powers et al., 2012,2017; Bauer et al., 2015; Urban et al., 2016; Alley et al., 2019).

Despite the improved achievements in the numerical models, the simulations and diagnosis in polar regions still entail much work. The first is the physics of the polar atmosphere; compared to the conditions of the middle-low latitude area, polar regions have special climate characteristics (air-sea-ice interaction), such as large ice sheets (glaciers), strong katabatic winds, strong sea-ice albedo feedback, and the role of clouds. Many of the key processes are not yet been well understood (Mayewski et al., 2009; Kennicutt et al., 2014, 2015, 2016). Second, the sparsity of the observational network is more severe in the Antarctic (Jung et al., 2015). For instance, the absence of long-term observations precludes definitive exploration of the large, climatically driven factors with only a few decades of observation. Additionally, due to maintenance challenges with the observing equipment to maintain good performance in harsh environments (Bromwich et al., 2013a; Jones and Lister, 2015), some observations lack consistency and continuity, and uncertainty is increased in the initial and lateral boundary conditions. Therefore, the polar-optimized model based on the Weather Research and Forecasting (WRF) has been developed and improved for polar applications by the Polar Meteorology Group (PMG) at the Byrd Polar and Climate Research Center of The Ohio State University. Previous versions of the model, known as "Polar WRF" (referred to as PWRF), have been used and evaluated in both the Arctic and the Antarctic. It is confirmed that PWRF is a powerful tool to investigate the weather and climate in these areas with good performance (Bromwich et al., 2009,2012,2013b,2018; Hines and Bromwich, 2008, 2017; Hines et al., 2011,2015,2019; Wilson et al., 2011,2012).

In recent work, based on the standard WRF 4.1.1, a new version of PWRF V4.1.1 (PWRF411) has been improved by PMG. This version of the model still serves as the mesoscale atmospheric model developed for both operational weather forecasting and research simulation. The model system is composed of a polar-optimized WRF, a WRF Pre-Processing System (WPS) and a WRF Data Assimilation System (WRFDA). Besides the characteristics of the previous version, the NoahMP Land Surface Model (LSM) was also improved for the first time in this version. Analogous to the development of the previous PWRF, a comprehensive evaluation of this new version is necessary. Currently, this model system version has been applied over the Arctic region by simulating polar lows near Svalbard during March 2013 and demonstrated good skill (Xue. et al., 2021). However, the Antarctic and Southern Ocean region that consists of a large ocean with limited observational data still need further evaluation. This article seeks to investigate the model performance, discusses the significant improvements of the new version model over the Antarctic and Southern Ocean including comparisons with the PWRF3.X. This study expands the investigation to all of Antarctica and Southern Ocean (Fig. 1) on an annual time scale and is the second in a series of verification studies documenting ongoing Polar WRF comprehensive evaluations in the Antarctic (Bromwich et al., 2013b). The impact of model improvements is assessed using different versions of Polar WRF (Polar WRF 4.1.1 and pioneer version 4.0.3) and standard WRF 4.1.1 for consistency with the reasonable verification data and methods of the previous studies (e.g., Hines and Bromwich, 2008; Bromwich et al., 2009, 2013b; Hines et al., 2011; Wilson et al., 2011), here three factors that are considered to influence model skills are focused on. That is model improvements, uncertainty in driving data, and intra annual variations in the large-scale atmospheric circulation. Assessment is mainly made using spatial correlation, bias, and root mean square error (RMSE) for a range of surface and upper air variables, using the station and sounding data for comparison. Also, the statistics focus on monthly and seasonal averages. The next sections are organized as follows. Section 2 describes the main improvement and configuration of Polar WRF 4.1.1. Section 3 depicts the experiments and

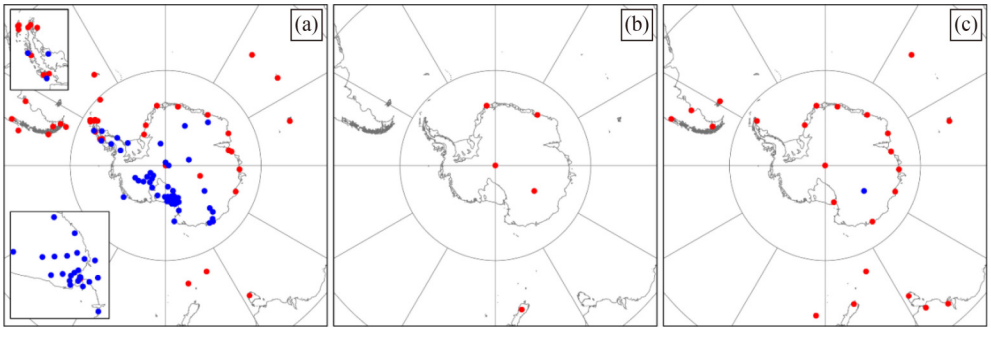

Fig. 1 Location of the observation sites. (a) Total surface data stations used (97). Red dots (43), blue dots (54) denotes stations from OGIMET and AMRC, respectively. Inset maps are for the Antarctic Peninsula (top) and Ross ice shelf (bottom). (b) Five BSRN stations used (DOM, GVN, LAU, SPO, and SYO). (c) Total upper air radiosounding stations used (22). Red dots (21), blue dots (1) denotes stations from UW and PNRA, respectively.

data used in this investigation. Section 4 gives the results of the benchmark simulations and sensitivity experiments for model versions, driving data and different simulation modes. Finally, a summary is provided in Section 5.

2 Polar WRF 4.1.1 (PWRF411)

The PWRF model system improved by PMG and built on previous success with the Pennsylvania State University (PSU)-National Center for Atmospheric Research (NCAR) Fifth-generation Mesoscale Model has been modified for use in polar regions (Polar MM5) (Guo et al., 2003; Bromwich et al., 2005). Based on WRF (ARW), PWRF has undergone releases from V3.0.1.1 (released November 2008) to V3.9.1 (released August 2017). The main target of the modifications is to promote the model skills in the polar regions. The key points include (a) optimize the treatment of heat transfer for ice sheets and revised surface energy balance calculation in the Land Surface Model (LSM) of Noah; (b) comprehensively describe sea ice in Noah; (c) improve cloud microphysics for polar regions. Many of these improvements are now part of the standard release with WRF (Hines and Bromwich, 2008,2017;Bromwich et al., 2009;Hines et al., 2011, 2015, 2019). At present, PWRF has attracted more than 480 registered users from 43 countries. For more details, please go to the associated website (available at Byrd Polar and Climate Research Center website).

Recently, PWRF has been upgraded to V4.1.1 based on the WRF version released in 2019; most of the optimizations were associated with the Noah and NoahMP LSM, with the representation of the heat transfer and fractional sea ice. Alternative specifications of the sea ice albedo (modified NoahMP minimum enabled setting of 0.8 over the ice sheets), allow users to specify spatially varying sea ice thickness (recommended: 1 m on the Antarctic) and snow depth (recommended max: 0.02−0.05 m, min: 0.002−0.02 m over the Antarctic) on sea ice. The droplet concentration in the Morrison 2-moment microphysics is reduced from 250 cm^{-3} to 50 cm^{-3}. WRF (and PWRF) uses fully compressible and non-hydrostatic equations for atmospheric dynamics on horizontal Arakawa C-grid staggering. The vertical coordinate is optional as either a terrain-following (TF) or hybrid vertical coordinate (HVC) hydrostatic pressure coordinate, and HVC is strongly recommended. In addition, to improve the adaptability in the polar regions, the freezing point of seawater is set at 271.36 K, the surface roughness over sea ice and permanent land ice is set at 0.001 m, the snow emissivity is set at 0.98, the snow density over sea ice is set at 300 kg/m^3, the thermal conductivity of the transition layer between the atmosphere and snow is taken as the snow thermal conductivity, and whenever the upper snow layer exceeds 20 cm depth, it is treated as if the snow were 20 cm thick for heat calculation; for an alternate calculation of the surface temperature over snow surfaces, the thermal diffusivity of the top 0.1-m-deep tundra soil is set to 0.25 $Wm^{-1}\cdot K^{-1}$, representative of highly organic soil (Bromwich et al., 2013b; Hines et al., 2015)

3 Experiments and data

3.1 Experiment strategy

To assess the performance of PWRF411 in the Antarctic and Southern Ocean, we use “Forecast Mode” and “Climate Mode” here referred to as FM and CM, respectively. FM is used to test the short-term forecast performance. The model is run in FM with a series of 48 h segments initialized daily at 0000 UTC with the first 24 h selected for model spin-up, which showed good performance in the previous studies (Wilson et al., 2011, 2012; Bromwich et al., 2013b; Hines et al., 2015). The CM for the long-term prediction skills with a series of 38–41-day initialized segments begins with 0000 UTC for 10 days before the 1st of each month. Using the first 10 days from each forecast allows for spin-up of the hydrological cycle and planetary boundary layer structure.

In this study, several downscaling simulations are required. There are two full-year experiments using PWRF411, and two seasonal experiments (JJA and DJF) with PWRF V4.0.3 (PWRF403) and WRF V4.1.1 (WRF411). The annual time scale was selected from June 2008 to May 2009, because the GPS radio occultation data will be used for this period to evaluate PWRF411 with data assimilation in the future. The full-year experiments used FM and CM simulations with different LSMs (Noah and NoahMP), while seasonal experiments with different LSMs were performed using FM simulations only. Moreover, ERA5 (ECMWF’s new atmospheric reanalysis) with pressure level data provides a detailed picture of the global weather and climate with higher horizontal resolution but less vertical resolution than ERA-Interim with model level data. We are also interested in whether ERA5 can improve the simulation or not. Thence, simulations performed with these lateral boundary conditions (ERA-Interim and ERA5) were investigated. To examine the influences of driving data, PWRF411 with FM driven by ERA5 Reanalysis (pressure level) in July 2008 and January 2009 with NoahMP and Noah were also conducted in this investigation. Therefore, a total of 10 sensitivity experiments were carried out. The summary of the characteristics for each experiment is given in Table 1.

3.2 Model configuration

The simulations use a single domain with 721 × 721 grid points on a polar stereographic projection (centered at the South Pole) and 15 km horizontal resolution, covering the Antarctic and most of the Southern Ocean (Fig. 1). In the vertical direction, the model uses the HVC hydrostatic

Table 1 Summary of the characteristics for each experiment

NO	Mode	Model	LSMs	Driven data	Simulation period	Comments (compared to benchmark experiment)
1	FM	PWRF411	NoahMP	ERA−Interim	Jun 2008−May 2009	benchmark experiment
2	FM	PWRF411	Noah	ERA−Interim	Jun 2008−May 2009	benchmark experiment
3	CM	PWRF411	NoahMP	ERA−Interim	Jun 2008−May 2009	different mode
4	CM	PWRF411	Noah	ERA−Interim	Jun 2008−May 2009	different mode
5	FM	PWRF403	NoahMP	ERA−Interim	Jun 2008−Aug 2008, Dec 2008−Feb 2009	different model version
6	FM	PWRF403	Noah	ERA−Interim	Jun 2008−Aug 2008, Dec 2008−Feb 2009	different model version
7	FM	WRF411	NoahMP	ERA−Interim	Jun 2008−Aug 2008, Dec 2008−Feb 2009	different model version
8	FM	WRF411	Noah	ERA−Interim	Jun 2008−Aug 2008, Dec 2008−Feb 2009	different model version
9	FM	PWRF411	NoahMP	ERA5	July 2008, Jan 2009	different driving data
10	FM	PWRF411	Noah	ERA5	July 2008, Jan 2009	different driving data

pressure coordinate with 71 model levels and a constant pressure surface at the top of 3 hPa, as it has been demonstrated that a higher model top provides better treatment of upward propagating gravity waves (Bromwich et al., 2005; Wilson et al., 2011). The lowest model level is ~−6 m, while the highest is ~4050 m, and the domain average is 225 m above ground level (AGL), with 26 levels below 850 hPa and 7−10 m level spacing in the lower 13 levels.

For ERA5 sensitivity experiments, initial and lateral boundary information is derived from ERA5 reanalysis every 3 h with 0.25° horizontal resolution and 37 standard pressure levels for the upper atmosphere. For others, the initial and lateral boundary conditions of simulations are interpolated from ERA-Interim reanalysis fields available every 6 h on 60 sigma levels and the surface at T255 resolution. It is the best representation of the atmospheric circulation near Antarctica among the reanalyses and is better represented in PWRF simulations (Bracegirdle and Marshall, 2012; Bromwich et al., 2013b; Hines et al., 2019). Nudging is used to reduce model drift (Miguez-Macho et al., 2004; Glisan et al., 2013; Hines et al., 2015; Bromwich et al., 2018). Based on a series of tests, spectral nudging is implemented on temperature, geopotential height, and wind above 200 hPa to improve model forecast skills. For instance, the bias of temperature at the model top level can be reduced from hundreds to ten of degrees C when CM simulations are run with nudging in July. However, the nudging of q vapor is not used because this step unreasonably reduces the precipitation. When nudging, coefficients for all three variables discussed above are set to 0.0003 and wave numbers selected to 17; as a result there are relatively small biases at the middle and upper vertical levels.

The physics parameterizations shown in Table 2 for this investigation rely upon a wide range of development and testing of PWRF over the polar regions (e.g., Hines and Bromwich, 2008, 2017; Bromwich et al., 2009, 2012, 2013b, 2018; Hines et al., 2011,2015, 2019; Wilson et al., 2011, 2012). For the microphysics parameterization, we used Morrison-2-mom, a scheme that has been extensively tested in the Arctic and Antarctic, and a liquid water droplet concentration specified 50 cm^{-3} is reasonable (Hines and Bromwich, 2017; Hines et al., 2019). For the PBL, the Mellor–Yamada–Nakanishi–Niino (MYNN) 2.5-level scheme (e.g., Nakanishi and Niino, 2009) is selected because of the biggest improvement of the reduction in the downward shortwave radiation bias through better cloud fraction and subgrid-scale mixing ratios of the new version WRF4.1.1 (Olson et al., 2019). The Kain-Fritch scheme (Kain, 2004) is used for cumulus parameterization, and the Rapid Radiative Transfer Model (RRTMG) (Iacono et al., 2008) is selected for longwave (LW) and shortwave (SW) radiation. Besides Noah LSM, the new NoahMP (e.g., Niu et al., 2011) with polar modifications that is different from previous works is utilized (e.g., Bromwich et al., 2013b).

3.3 Validation data and methods

The observation sites for surface analysis of 2 m temperature, 2 m dew point, surface pressure, and 10 m wind speed are given in Fig. 1(a) (97 stations, and some of them are manual stations). There are two data sources. One is 3 hourly quality-controlled automatic weather station (AWS) data provided by the Antarctic Meteorological Research Center (AMRC) at the University of Wisconsin-Madison (available at AMRC and AWS website). Second, because most of the AWSs from AMRC are deployed in the Antarctic inland area, for the coastal and Southern Ocean region, data from 43 other main stations were obtained from the OGIMET database that collects and processes freely available data, mainly from the National Oceanic and Atmospheric Administration (NOAA) (available at OGIMET website). For AWS, as the wind speed measured at ~3 m, we adjusted this speed using the logarithmic wind profile up to 10 m. The surface downwelling shortwave and

Table 2 Main setup of experiment strategy and model configuration

Description	Forecast mode (FM)	Climate mode (CM)
Horizontal resolution	15 km	
Simulation	Short-term48 h run	Long-term38–41days run
Spin-up	24 h	10 days
Lateral boundary data	ERA-Interim/ERA5	
Vertical Level	71 levels, Model top level at 3hPa	
Coordinate	Hybrid Vertical Coordinate, eta = 0.3	
Land surface options	Noah (2) NoahMP (4)	
Microphysics	Morrison 2-mom (10)	
PBL scheme	MYNN2 (5)	
Short/Long wave	Both RRTMG (4&4)	
Cumulus	Kain-Fritsch (1)	
Surface layer	MYNN (5)	
Nudging	Wave number 17, spectral nudging t, ph, u,v, above 200 hPa	
Seaice options	seaice_thickness_default = 1.0 seaice_albedo_default = 0.80 seaice_snowdepth_max = 0.02 (min)−0.05 (max) Change with monthly time from 0.02 in July and 0.05 in January with 0.005 interval for each monthy with annual cycle. seaice_snowdepth_min = 0.002 (min)−0.02 (max) Change with monthly time from 0.02 in July and 0.002 in January with 0.003 interval for each monthy with annual cycle.	

longwave radiation analysis uses Baseline Surface Radiation Network (BSRN) data, which has only 5 stations in the domain (DOM, GVN, LAU, SPO, and SYO, Fig. 1(b)).

Upper air analysis is conducted using the atmospheric sounding date from University of Wyoming (UW, available at University of Wyoming website) and the Antarctic Meteo-Climatological Observatory funded by the Italian National Program of Antarctic Research (PNRA, available at Antarctic Meteo-Climatological Observatory website); both include the data at all levels (Fig. 1(c)). Here, the analysis was performed on the following levels: 975 hPa, 950 hPa, 850 hPa, 700 hPa, 500 hPa, 300 hPa, 200 hPa, 150 hPa, and 100 hPa, where observations and model results are interpolated to the same pressure levels.

Because of the extreme climate and challenging environment, there are many missing or unusable measured data (e.g., Bromwich et al., 2013a; Jones and Lister, 2015). Therefore, data quality control is necessary. A visual examination of individual time series for potentially erroneous characteristics is performed for each station and variable. These include, for example, large spikes in surface pressure, unreasonable continuous zero values or negative number wind speed reading with non-missing data, etc. All these singular points are filled with missing values. Finally, the number of stations for each variable used may not be the same during different months.

Additionally, ERA-Interim reanalysis and forecast are used for complementary comparison. For the ERA-Interim forecast, data come from twice daily forecasts at 00/12 UTC and up to the range of 12 h, respectively (Berrisford et al., 2011). Here, the 3 hourly variables from the forecast data are extracted and interpolated onto the domain by the WPS.

4 Results

4.1 Benchmark experiment

In this investigation, PWRF411 with FM simulations from June 2008 to May 2009 employing different LSMs is used as the benchmark experiment to evaluate the model performance (hereafter NoahMP and Noah). The comprehensive evaluation here draws on the methods Bromwich et al. (2013b) used for the earlier versions of Polar WRF 3.X (here referred to as B13 below). As the focus is on the new model version, especially for the PWRF with NoahMP that is first studied here in the Antarctic, both the performance of NoahMP and Noah will be discussed in detail and we have elected to provide limited comparisons with B13.

4.1.1 Surface variable evaluation

The 2 m temperature and dew point results of the model are adjusted to the observation station height using the environmental lapse rate of 6.5 K/km. Likewise, the surface pressure is corrected hypsometrically to station elevations with the temperature adjusted at each observation time (Wilson et al., 2011). The means of model correlation, bias, and RMSE are computed with respect to the 3 hourly observations for the 2 m air temperature, 2 m dew point, surface pressure, 10 m wind speed, longwave and shortwave down. The stations used

for monthly statistics have missing values < 50%.

4.1.1.1 2 m temperature and dew point

Figure 2 shows the model skills in forecasting the temporal variability of the 2 m air temperature. The annual average correlations exceed 0.8 and are lower in the summer time. In January, the lowest value of ~0.73 indicates difficulties in forecasting temporal variability at the peak of the diurnal cycle (Bromwich et al., 2013b). Bias for the whole domain is generally cold. The annual average bias of NoahMP is −0.98°C and Noah is −0.72°C, both absolute values are less than 1°C. The spatial patterns of the statistics (not shown) demonstrate slightly warmer areas situated inland and much colder areas at coastal stations. Although NoahMP has a larger cold bias than Noah (the reason is given in Section 4.2.3) but has a better correlation and RMSE. With further analysis of the diurnal variations, NoahMP shows better performance than Noah but still exhibits a too strong diurnal cycle in the coastline. More details will be discussed below in Section 4.2.3.

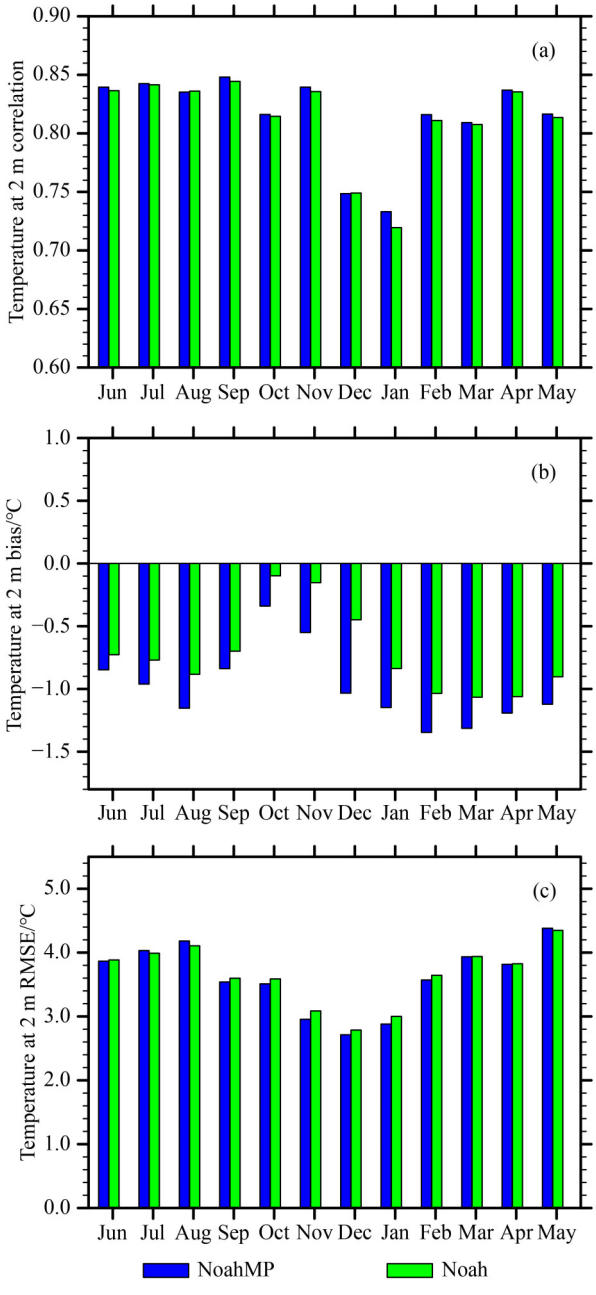

Fig. 2 (a)–(c) Domain-averaged monthly 2 m air temperature statistics, total station numbers from June 2008 to May 2009 are 70, 67, 67, 67, 67, 75, 87, 87, 91, 90, 89, 85, respectively

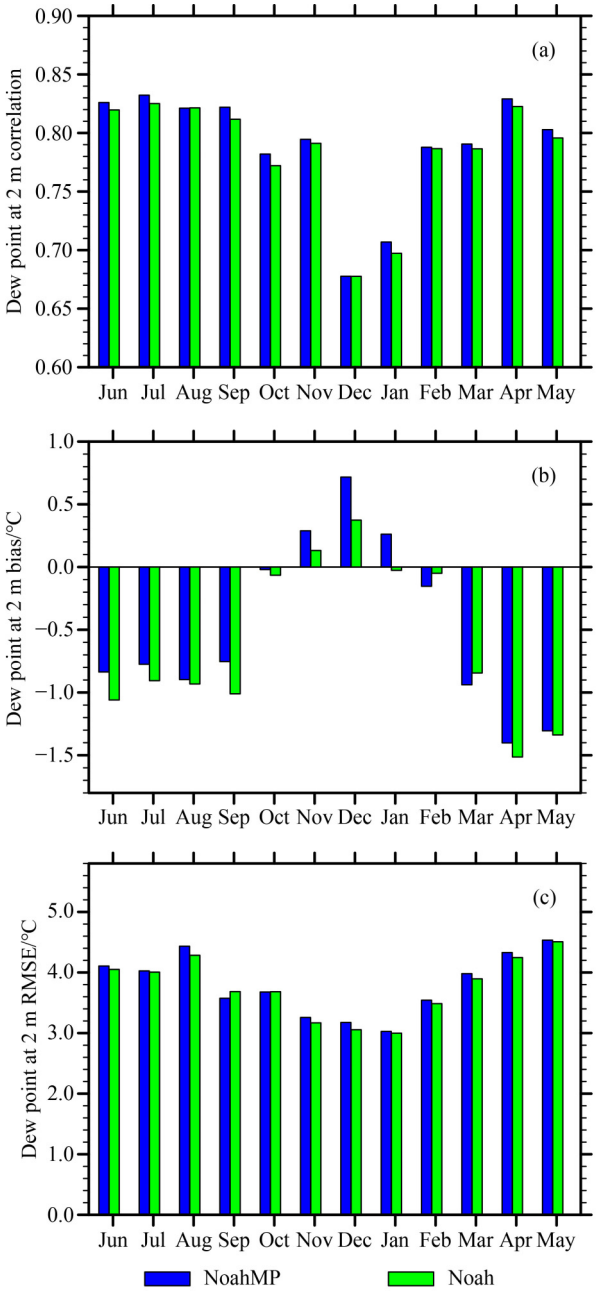

Fig. 3 (a)–(c) Domain-averaged monthly 2 m dew point statistics, total station numbers from June 2008 to May 2009 are 58, 57, 57, 57, 57, 64, 71, 70, 75, 73, 73, 70, respectively

For the 2 m dew point, the domain-averaged monthly correlation, bias, and RMSE are plotted in Fig. 3. The annual average correlation is ~0.78. It is similar to the 2 m air temperature, which is lower in summer and higher in winter. The annual average bias of NoahMP is ~−0.48 while Noah is ~−0.6°C. Both of them have positive values (the peak at December) in most of the summertime but negative during other seasons. And, NoahMP shows higher correlations and smaller bias except during summertime than Noah.

Compared to the 2 m temperature and dew point with B13, the correlations generally increase by more than 0.1. The absolute values of bias and RMSE decreased ~1°C−1.5°C, respectively. It indicates that PWRF 411 shows better performance than the earlier versions of PWRF3.X.

4.1.1.2 Surface pressure and wind speed

PWRF411 exhibits excellent performance in surface pressure. During the annual cycle, the correlation is ~0.99 and illustrates positive bias, as shown in Fig. 4. Both the bias and RMSE have a strong seasonal cycle with minima (~0.5 hPa, 1.5 hPa) at the peak of summer (January). The spatial analysis found that larger biases occurred in the steep mountain area and indicates that challenges remain in complex terrain areas (such as the Transantarctic Mountains) for the surface pressure forecast. NoahMP and Noah show similar results, except the former has a slightly higher bias during summer, which is most likely caused by the colder domain-averaged surface temperature mentioned in Section 4.1.1.1

The annual average correlation of 10 m wind speed exceeded 0.65, with no well-defined difference between winter and summer (Fig. 5). Values from NoahMP and Noah show slight differences for most months. Unlike the correlation, the bias and RMSE show significant seasonal variation with the maximum positive bias in July (NoahMP is ~0.98 m/s, Noah is ~1.2 m/s) and minimum bias in January (both < 0.1 m/s). This aspect is likely related to seasonal variation in absolute wind speeds that higher speeds produce larger biases (Bromwich et al., 2013b). Most of the time, NoahMP has a better bias and RMSE than Noah (Fig. 5).

For the surface pressure and wind speed, it also indicates improvement compared with B13. Specifically, the higher correlation and lower bias suggest PWRF411 improves skills in the summertime. We believe this benefits from the optimization of the reduction in downward shortwave radiation in WRF411 and the modified NoahMP of PWRF 411.

4.1.1.3 Downwelling surface longwave and shortwave radiation

Five BSRN stations at DomeC (DOM), Amundsen-Scott (SPO), Neumayer (GVN), Lauder (LAU), and Syowa (SYO) in this domain are used to examine the surface

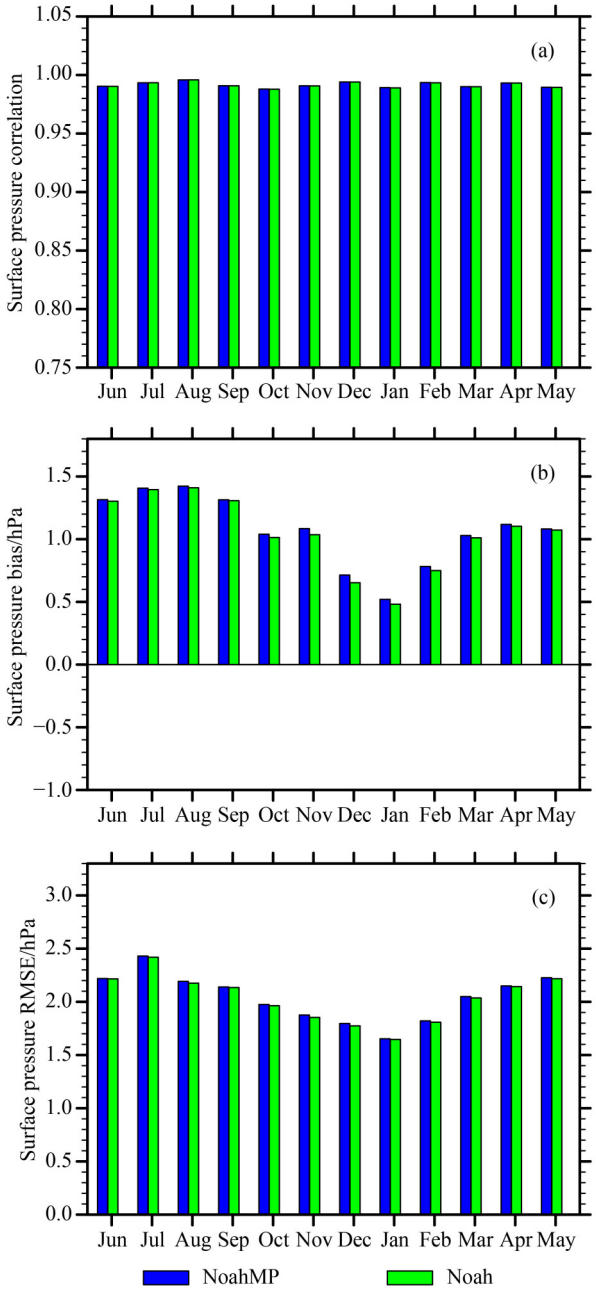

Fig. 4 (a)–(c) Domain-averaged monthly surface pressure statistics, total station numbers from June 2008 to May 2009 are 67, 66, 62, 66, 64, 71, 84, 83, 84, 83, 83, 80, respectively.

radiation balance. Similar to evaluation of the variables above, the correlation, bias, and RMSE were calculated monthly for each station (not shown). For longwave down (LWD), inland sites (DOM, SPO) have a positive bias, whereas coastal sites (GVN, LAU, and SYO) show negative values throughout the year. Correspondingly, shortwave down (SWD) demonstrates positive bias at coastal sites, but differences emerge between DOM and SPO. DOM, the location on the East Antarctic Plateau, shows negative bias most of the time. In contrast to DOM, there is generally a positive bias at SPO. For

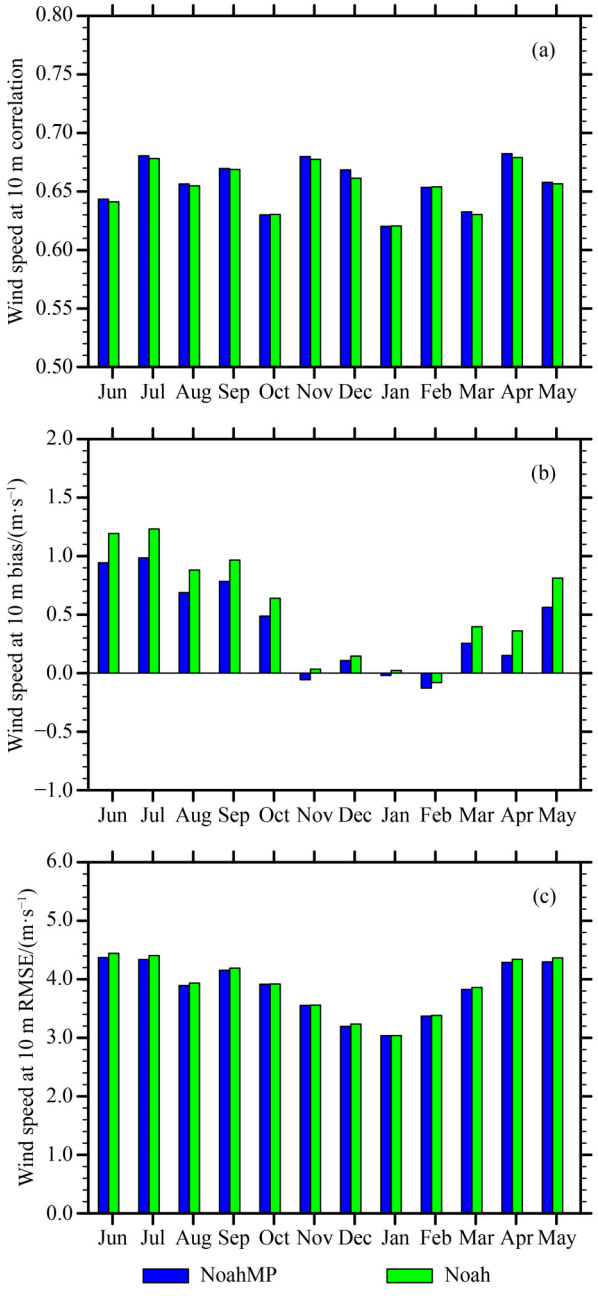

Fig. 5 (a)–(c) Domain-averaged monthly 10m wind speed statistics, total station numbers from June 2008 to May 2009 are 59, 59, 56, 59, 58, 65, 80, 80, 83, 84, 78, 71, respectively

clarity, seasonal (monthly averaged JJA, SON, DJF, and MAM) and annual (AAN, average 12 months) statistics are also carried out for LWD and SWD. As NoahMP and Noah show the same result, only the results of NoahMP are given in Table 3. In summer (DJF), the correlation and RMSE of LWD are lower than those of other seasons, except for LAU. The bias for GVN, SYO, and LAU is larger in summer. Compared with LWD, SWD usually has a higher correlation related to the diurnal cycle. The largest bias and RMSE still appear in summer.

To assess the representativeness of BSRN stations, 13 additional coastal locations (Belgrano2, Casey, Davis, Dumont D'Urville, Halley, King Sejong, Mawson, McMurdo, Mirny, Evangelistas, Hobart, Mount Siple and and Possession Island) and 10 inland stations (Vostok, Baldrick, D-85, Henry, Harry, JASE2007, Kominko-Slade, Nico, PANDA-South, and Theresa) are selected in the domain. The model forecast data interpolated to these sites and monthly means are calculated with different classified sites. Figures 6(a) and 6(b) shows the monthly mean of LWD. "BSRN" and "Model" are the averages of the observations and forecasts, respectively. "Group" corresponds to additional locations with similar characteristics from the forecast (including BSRN stations). LWD shows a seasonal cycle with a maximum in summer. At the coastal sites, the forecast average, whether it is from BSRN stations or additional stations, is generally lower than the average observation at GVN, LAU, and SYO. Conversely, the forecast average is always higher at the interior land locations. Although the same additional stations were used to check SWD, the inland stations were divided into two groups based on the previous results. One part is the 4 site-like characteristics of DOM on the East Antarctic Plateau (Vostok, D-85, JASE2007, and PANDA-South). The other is around the South Pole and the West Antarctic (Baldrick, Henry, Harry, Kominko-Slade, Nico, Theresa). In Figs. 7(a)–7(c), SWD demonstrates a much stronger seasonal cycle than LWD and reaches a peak in December when the sun moves to the southernmost latitude of the southern hemisphere. Stations located on the coastline, South Pole and West Antarctic show more SWD than the observations. Different results occur for the East Antarctic Plateau. From November to February, the forecast displays insufficient SWD and reverses at other times (except June, July, and May). The spatial pattern of LWD and SWD discussed above is very likely the cause of the forecast appearing much colder in the coastline and slightly warmer over interior land.

4.1.2 Upper air variable evaluation

To describe the PWRF upper air performance in detail, monthly mean vertical profiles of temperature, wind speed, height, and relative humidity are calculated using 0000 UTC and 1200 UTC soundings with matching model output times. A total of 22 atmospheric sounding stations are used, which are distributed in the Antarctic inland (2 stations: Amundsen-Scott, Dome C), coast (10 stations: Casey, Davis, Dumont D'Urville, Halley, Mawson, McMurdo, Mirnyj, Neumayer, Novolazarevskaja, and Syowa) and the Southern Ocean (10 stations: Adelaide, Comodoro Rivadavia Aero, Hobart Airport, Kerguelen, Macquarie, Marion, Melbourne Airport, Mount Pleasant, Puerto Montt and Punta Arenas) (Fig. 1(c)). The observations include full pressure levels (containing the

Table 3 Seasonal and annual statistics of downwelling surface longwave and shortwave radiation statistics for the five BSRN stations. Here, only the results of NoahMP with FM mode are shown

	Longwave				AAN	Shortwave				AAN
	JJA	SON	DJF	MAM		JJA	SON	DJF	MAM	
Correlation										
DOM	0.81	0.83	0.61	0.84	0.77	–	0.99	0.99	0.98	0.99
SPO	0.88	0.88	0.58	0.84	0.79	–	0.95	0.75	0.97	0.87
GVN	0.86	0.84	0.77	0.83	0.83	0.72	0.96	0.98	0.91	0.91
LAU	0.70	0.80	0.81	0.70	0.75	0.85	0.92	0.90	0.90	0.89
SYO	0.82	0.84	0.70	0.82	0.79	0.86	0.99	0.96	0.92	0.93
Bias (W/m^2)										
DOM	14.56	20.77	15.83	16.78	16.99	–	−1.83	−12.49	−9.99	−7.64
SPO	11.04	11.11	5.61	10.89	9.66	–	16.42	19.68	5.29	16.23
GVN	−10.83	−11.00	−17.38	−9.66	−12.22	−1.47	23.11	28.30	9.09	16.23
LAU	−9.19	−14.60	−15.18	−14.99	−13.49	15.65	43.30	60.48	25.77	36.30
SYO	−9.20	−4.60	−17.02	−10.18	−10.25	0.79	16.17	38.72	9.25	17.78
RMSE (W/m^2)										
DOM	19.20	25.02	26.59	19.61	22.61	–	20.40	34.87	14.71	21.79
SPO	17.71	17.98	21.95	19.98	19.40	–	22.54	38.58	9.70	27.58
GVN	24.93	26.87	31.66	25.25	27.18	10.15	54.91	60.20	27.32	40.69
LAU	27.31	25.49	24.79	28.79	26.59	67.43	117.47	152.05	87.00	105.99
SYO	26.90	23.46	29.78	26.00	26.54	9.38	33.64	79.54	27.79	40.81

major standard pressure levels), and model profiles consist of all 71 vertical levels up to the top of 3 hPa. Therefore, the results for all variables are vertically interpolated from 1000 hPa to 25 hPa with 25 hPa intervals to pursue additional details and contain the standard vertical levels. Only the results from NoahMP are shown because slight variations occur between NoahMP and Noah in the upper air statistics (not shown). Because the upper air variable differences are generally smaller than that at the surface, July and January are used to represent winter and summer, respectively, to assess the model performance with the largest seasonal contrast. The domain-averaged statistics for all near-surface variables and the other standard levels are presented in Table 4. In general, the results show better performance than B13 especially in higher levels (above ~500hPa). This suggests that the higher vertical resolution (71 levels) and model top (3 hPa) along with nudging is reasonable and highly recommended.

Figure 8 exhibits the monthly mean vertical temperature profiles for 22 stations and the domain-averaged results for July 2008 and January 2009. Because of complex Antarctic topography and many missing observations above 100 hPa, profiles are plotted at 975 hPa (Amundsen-Scott and DomeC at 650 hPa) up to 100 hPa with the standard deviation at each standard level (only for the domain-averaged profiles, see Figs. 8(c) and 8(f). The model depicts the characteristics of the temperature profile accurately in both months (Figs. 8(a)–8(f)). For all stations, there is a larger temperature spread (~30°C) near the surface and the top. Both observation and forecast clearly show the tropopause near 300 hPa, even the details that some stations in the Southern Ocean are slightly higher because they are warmer. For temperature statistics in Table 4, the correlations range from 0.83 to 0.98, which is lower at the bottom and top levels. Unlike the surface bias and RMSE discussed above, the bias and RMSE of the upper air temperature are very small, and the maxima are less than 1°C and 2°C, respectively. Moreover, biases below 925 hPa show that they are cold near the surface, which is consistent with previous conclusions. RMSE values that are large at the bottom and top levels indicate that temperatures are relatively variable, as supported by the individual station profiles, which show a larger temperature spread in Fig. 8.

Similar to the temperature, the monthly mean vertical profiles of horizontal wind speed match the same stations for July 2008 and January 2009, as shown in Fig. 9. The overall profiles are similar between observations and forecasts in both months. Below the tropopause (~300 hPa), wind speeds generally increase, whereas they decrease and increase again above 200 hPa in July but only decrease above 300 hPa in January. For most stations, there is a larger spread at the top, and the maximum wind speed appears near 300 hPa or a little higher for stations sited in the Southern Ocean, which have slightly higher tropopauses. The correlations of July are slightly higher than those of January, and both increase upward from the

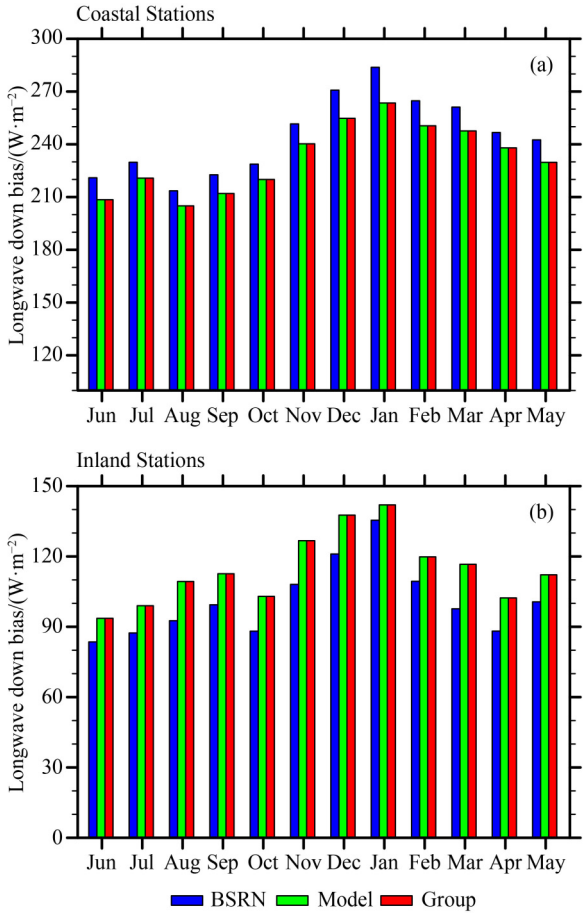

Fig. 6 (a)–(b) Mean monthly incident longwave radiation at the surface. BSRN, Model, denotes station-averaged observation and forecast data (NoahMP FM simulation) respectively. Group is computed for additional locations of similar characteristics from the forecasts. (a) Coastal stations, Group number is 16, including 3 BSRN station sites (GVN, LAU, SYO). (b) Inland stations, Group number is 12, including 2 BSRN station sites (DOM, SPO).

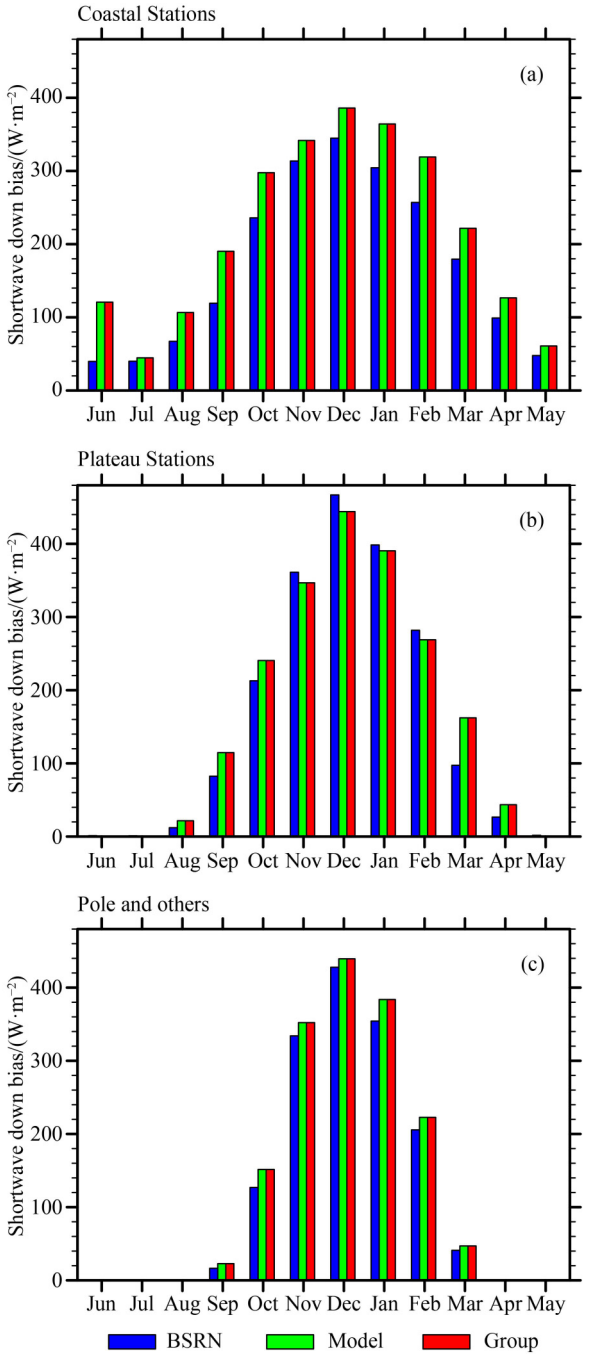

Fig. 7 (a)–(c) Mean monthly incident shortwave radiation at the surface. BSRN, Model, denotes station-averaged observation and forecast data (NoahMP FM simulation) respectively. Group is computed for additional locations of similar characteristics from the forecasts. (a) Coastal stations. Group number is 16, including 3 BSRN station sites (GVN, LAU, SYO). (b) Plateau stations. Group number is 5, including 1 BSRN station site (DOM). (c) South Pole other inland stations. Group number is 7, including 1 BSRN station site (SPO).

surface, reaching peaks at 300 hPa. The model wind speeds near the surface and top are slightly stronger but weaker at other levels with positive and negative biases, respectively (Table 4). Furthermore, the positive bias below 925 hPa is consistent with the result of a slightly stronger 10 m wind speed as revealed earlier.

Table 4 also exhibits statistics of geopotential height and relative humidity. For height, the correlations are always strong and remain above 0.9 in both months. The bias and RMSE have large values above 300 hPa, indicating that greater variability occurs near the tropopause. Similarly, the biases (absolute value) and RMSE in July usually exceed those in January, indicating more variability in July, except at the surface. Relative humidity statistics show that forecasting is still facing challenges. Because measurements are less reliable in the upper atmosphere, only values for levels below 300 hPa (throughout the troposphere) are given. Correlations below 600 hPa increase with height but decrease above that. Biases in July below 600 hPa show a negative value (except 975 hPa) but are positive in January. In addition, positive biases at 975 hPa in January are consistent with the result of a higher 2 m dew point.

Table 4 Domain-averaged upper air statistics for July 2008 and January 2009

Level (hPa)	July			January			July			January		
	CORR	BIAS	RMSE	CORR	BIAS	RMSE	CORR	BIAS	RMSE	CORR	BIAS	RMSE
Air temperature (°C)							**Height (m)**					
975	0.83	−0.85	2.18	0.83	−0.10	1.64	0.99	−2.46	8.45	0.94	−2.28	11.05
950	0.90	−0.28	1.79	0.88	−0.50	1.56	0.99	−0.69	9.88	0.96	−1.43	10.39
925	0.94	0.05	1.47	0.91	−0.28	1.22	0.99	0.70	9.43	0.98	0.19	9.46
850	0.95	0.12	1.26	0.94	−0.20	1.03	0.91	−1.08	25.39	0.98	0.08	9.56
700	0.96	−0.13	1.00	0.96	−0.19	0.93	0.99	1.95	11.27	0.98	−0.13	10.26
600	0.97	0.04	0.99	0.96	0.00	0.90	0.98	1.21	16.38	0.97	−1.06	15.34
500	0.97	0.09	0.79	0.98	−0.02	0.74	0.99	5.74	15.08	0.99	2.36	13.51
400	0.98	0.07	0.69	0.98	−0.04	0.70	0.99	4.16	15.36	0.99	0.88	14.71
300	0.96	0.00	0.81	0.96	0.01	0.87	0.97	6.96	27.36	0.99	2.01	17.53
200	0.96	0.20	1.15	0.97	0.01	1.04	0.97	17.97	30.01	0.97	11.90	29.46
150	0.95	0.10	0.98	0.95	−0.15	0.93	0.96	13.25	33.66	0.96	4.98	28.13
100	0.91	0.11	1.04	0.92	−0.25	1.04	0.95	15.88	45.46	0.96	4.33	24.58
Wind speed (m/s)							**Relative humidity (%)**					
975	0.74	0.73	3.36	0.66	0.31	3.03	0.52	3.94	19.06	0.51	7.90	17.03
950	0.83	0.16	3.93	0.75	0.24	3.34	0.64	−2.09	17.12	0.66	5.12	14.59
925	0.86	−0.10	3.76	0.84	−0.13	3.24	0.73	−3.21	16.90	0.72	4.37	14.59
850	0.91	−0.41	3.09	0.87	−0.12	2.76	0.82	−2.55	17.07	0.75	3.11	15.68
700	0.92	−0.39	2.53	0.91	−0.49	2.19	0.83	−0.89	17.16	0.79	1.51	17.26
600	0.93	−0.07	2.77	0.92	−0.30	2.25	0.82	1.23	16.52	0.79	2.32	17.12
500	0.95	−0.22	3.07	0.95	−0.23	2.22	0.80	1.63	16.64	0.75	4.75	17.45
400	0.96	−0.32	3.41	0.95	−0.67	3.47	0.78	2.60	15.26	0.73	6.54	16.72
300	0.97	−0.17	3.54	0.97	−0.23	2.95	0.77	4.22	13.55	0.75	7.81	16.46
200	0.97	0.34	2.90	0.97	0.04	2.27						
150	0.97	0.53	2.67	0.96	0.15	2.16						
100	0.94	0.36	2.96	0.88	0.34	2.59						

Note: Bold represents standard levels

4.2 Sensitivity to different model versions, driving data and simulation modes

This section discusses experiments with different model versions, driving data and stimulation modes stated in Section 3.1. The observation data and surface variable evaluation methods used are as same as the benchmark experiments.

4.2.1 Comparison of different versions

Upon the release of new WRF 4.X versions, PMG upgraded PWRF to V4.0.3 and V4.1.1 (PWRF403 and PWRF411, respectively). The most significant new features of the two versions are modified NoahMP (setting minimum albedo over ice sheets), and the improvement in the downward shortwave radiation implemented in WRF411. Thence, the 2 m temperature, LWD, and SWD are mainly used for evaluation.

Domain-averaged 2 m temperature statistics for January and July are shown in Table 5. Due to the setting minimum albedo being 0.8 over the ice sheet, PWRF411 with NoahMP exhibits an obvious difference from the other models in January. It shows a larger cold bias but has a better correlation and RMSE. In fact, the spatial patterns of the biases could reveal the reason. The biases of each station versus observation for PWRF411 NoahMP, PWRF411 Noah and WRF NoahMP are plotted in Fig. 10. Stations sited in the Southern Ocean and the Antarctic coastal areas did not show significant differences but exhibited a contrast in the interior. For WRF NoahMP, most stations located inland show the largest warm bias (>4°C or more, see Fig. 10(c)). Although PWRF411 NoahMP and Noah have a better result, there are differences over the ice sheet, especially near the South Pole and Ross Ice Shelf, where the former has a smaller bias. These characteristics account for why PWRF411 NoahMP shows the coldest (−1.15°C) domain-

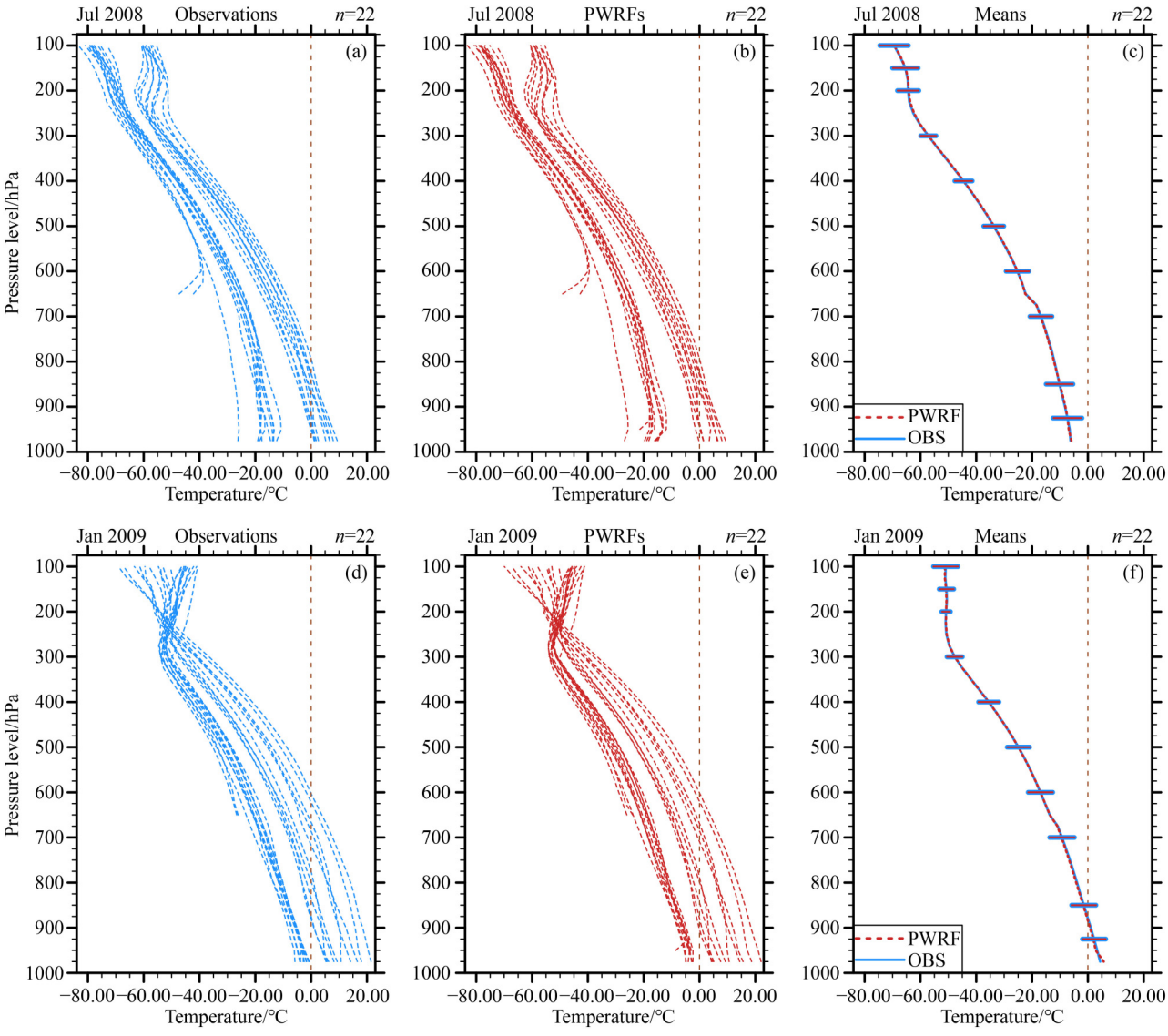

Fig. 8 (a)–(c) July 2008 and (d)–(f) January 2009 monthly means vertical temperature profiles for 22 stations of observations (Figs. 8(a) and 8(d)), PWRFs (Figs. 8(b) and 8(e)), and means (Figs. 8(c) and 8(f)) plotted from 925 hPh to 100 hPa vertical levels. Figures 8(a), 8(b), 8(d), and 8(e) are individual stations plotted with dashed lines. Figures 8(c) and 8(f) are the station means with the blue solid line for observation and red dashed line for PWRF. Thick horizontal bars (blue) and thin horizontal bars (red) at the vertical standard levels represent standard deviation for observation and PWRF, respectively.

averaged bias and WRF411 NoahMP has the warmest (0.74°C) in Table 5.

For more details, the bias spatial patterns of the model versus ERA - Interim reanalysis and forecast were also analyzed. As the two have very consistent results that predicted near-surface temperatures agree well spatially with observations, only the model versus ERA - Interim forecast is given in Fig. 11. The results also support that WRF NoahMP is much warmer than PWRF and modified NoahMP in PWRF411 shows better forecast skill for the 2 m temperature in the Antarctic.

In general, a higher surface temperature is related to an excessive LWD, and vice versa. As there are only five BSRN stations in this domain, unlike the 2 m temperature, only the bias spatial patterns (here model versus the ERA-Interim forecast) for PWRF411 NoahMP, PWRF411 Noah and WRF NoahMP are plotted. The spatial patterns are similar in July but different in January. In Fig. 12, a larger positive bias means that WRF NoahMP without modification generates too much LWD over the Antarctic. Furthermore, as discussed in Section 4.1.1.3, all of the models underestimate LWD in the coastal area but exhibit excess around the Southern Ocean.

Finally, six model bias spatial patterns (model versus ERA - Interim forecast) for SWD in January are shown in Fig. 13. First, both PWR411 and WRF411 clearly reduce the bias of SWD, especially over the ocean. Then, PWRFV411 and PWRFV403 also show better negative bias in the scope of 60°E−60°W, 30°−40°S in the

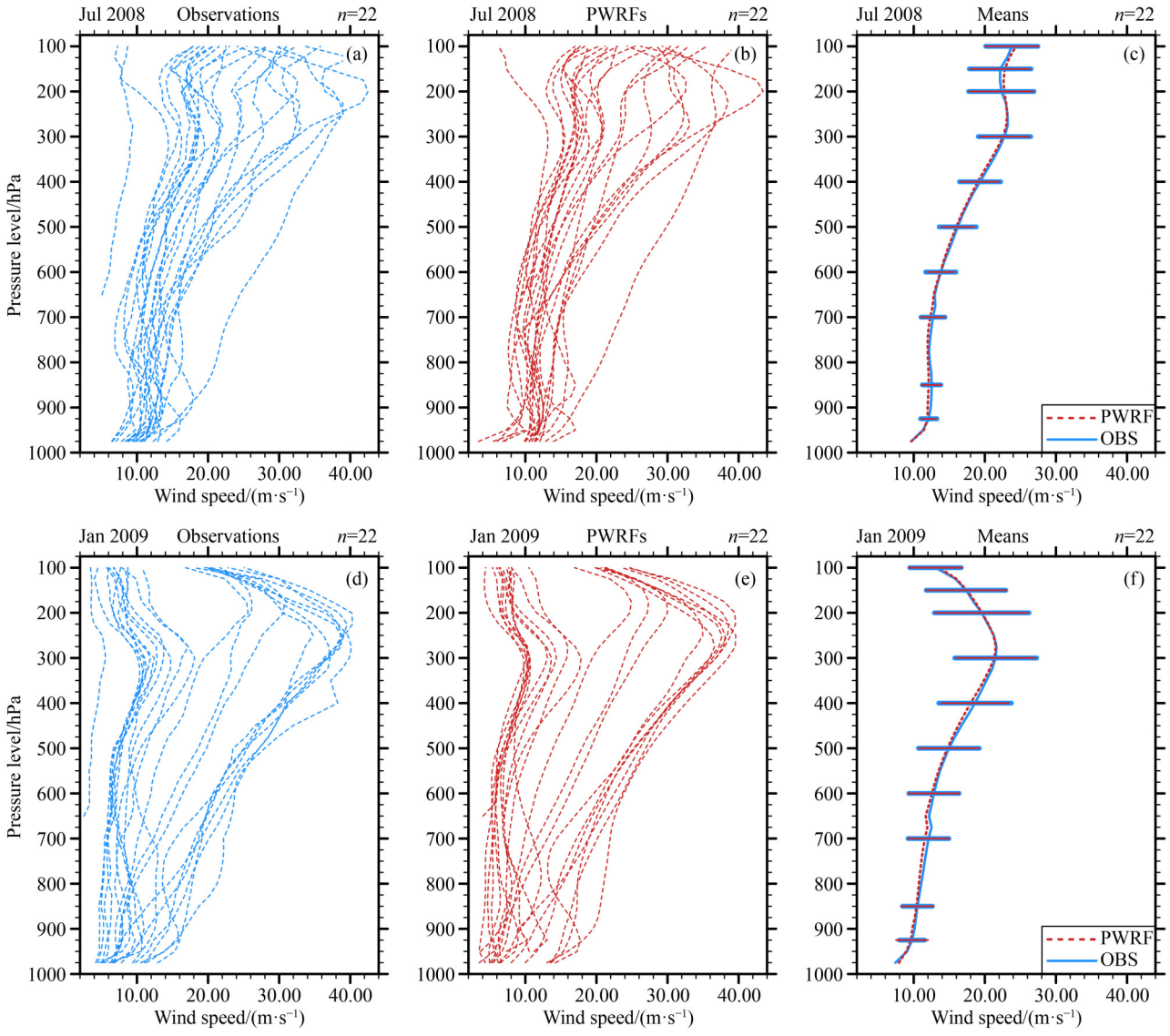

Fig. 9 (a)–(c) July 2009 and (d)–(f) January 2009 monthly means vertical distribution of horizontal wind speed profiles for 22 stations of observations (Figs. 9(a) and 9(d)), PWRFs (Figs. 9(b) and 9(e)), and means (Figs. 9(c) and 9(f)) plotted from 925 hPh to 100 hPa vertical level. Figures 9(a), 9(b), 9(d), and 9(e) are individual stations plotted with dashed lines. Figures 9(c) and 9(f) are the stations means with the blue solid line for observation and red dashed line for PWRF. Thick horizontal bars (blue) and thin horizontal bars (red) at the vertical standard levels represent standard deviation for observation and PWRF, respectively.

Southern Ocean. Besides that, for NoahMP, PWRF411 improves the insufficient SWD, which illustrates the maximum negative value on the East Antarctic Plateau. In addition, the bias spatial patterns of PWRF411 NoahMP still display slight negative biases on the East Antarctic Plateau and are inverted in other areas, such as the West Antarctic (Fig. 13(a)). This phenomenon is consistent with the findings in Section 4.1.1.3.

4.2.2 Difference initial and lateral boundary conditions

To assess the influences of initial and lateral boundary conditions, FM simulations with the different LSMs (NoahMP and Noah) were examined by driving with ERA5 (E5) and ERA-Interim (EI) reanalyses. Here, the January and July correlations (CORR), RMSE, and bias from reanalysis (E5, EI) and forecasts (FM, 48 h cold start runs with spin-up for the first 24 hour) comparison with observations are given in Table 6. Generally speaking, all the forecasts show reasonable results compared to the reanalysis. The performance of forecasts driven by ERA5 was comparable to or did not show obvious improvement than ERA-Interim. For the 2 m temperature, the domain-averaged statistics from the forecasts driven by ERA5 show a larger negative bias (both NoahMP and Noah). This is probably because the model driven by ERA5 improved the warm bias near the South Pole and East Antarctic Plateau but still had large cold biases at the coastal stations, from further spatial analysis (not shown). Comparing all reanalysis and forecast

Table 5 Domin-averaged 2 m temperature statistics for January and July (°C, *n* = number of stations)

Model	PWRF403		WRF411		PWRF411	
LSM	**NoahMP**	**Noah**	**NoahMP**	**Noah**	**NoahMP**	**Noah**
January (*n* = 87)						
CORR	0.69	0.71	0.69	0.72	0.73	0.72
BIAS	0.62	−0.95	0.74	−0.86	−1.15	−0.84
RMSE	3.27	3.06	3.29	3.06	2.88	3.00
July (*n* = 67)						
CORR	0.85	0.85	0.84	0.84	0.84	0.84
BIAS	−0.87	−0.72	−0.96	−0.77	−0.96	−0.77
RMSE	4.08	4.03	4.03	3.99	4.03	3.99

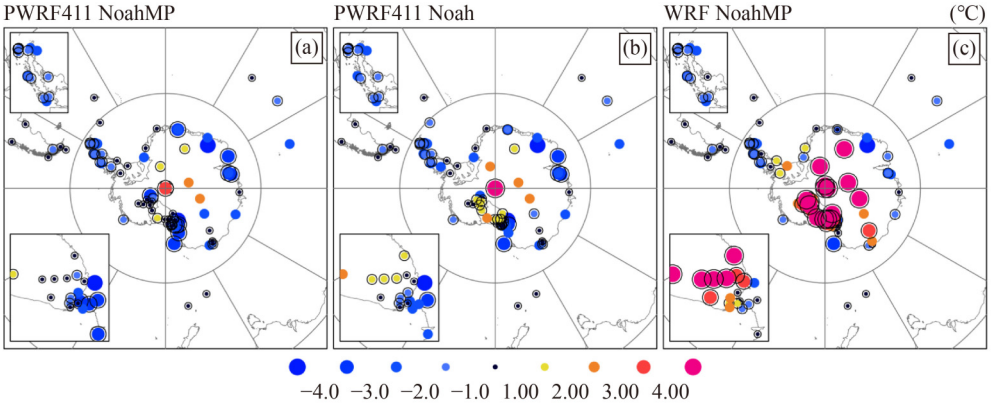

Fig. 10 The bias (Model versus Observations) of 2 m temperature for (a) PWRF411 NoahMP, (b) PWRF411 Noah, and (c) WRF NoahMP in January 2009. The total station number is 87. Dots with a black circle means the correlation ⩾ the annual average of domain-averaged (0.81).

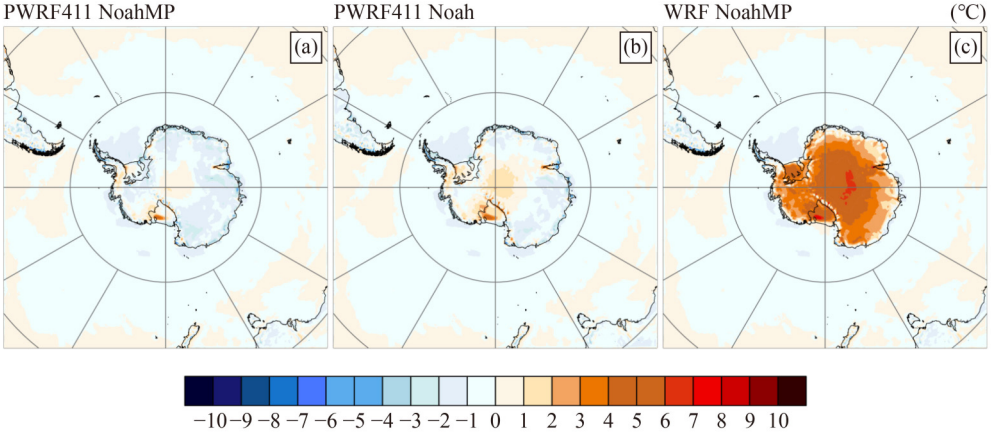

Fig. 11 The bias of (Model versus ERA-Interim forecast) of 2 m temperature for (a) PWRF411 NoahMP, (b) PWRF411 Noah, and (c) WRF NoahMP in January 2009.

results, surface pressure shows ERA-Interim has better performance than ERA5. In addition, the lower correlations (0.97), bigger bias and RMSE of ERA5 with observations were caused by some high altitude AWS (such as DC2, JAS, HEN) in January. The 2 m Dew point and 10 m Wind speed still face challenges in the forecast, however, ERA5 shows better correlation and RMSE more often than ERA-Interim.

Based on the findings discussed above, some possible reasons were inferred as follows. First, ERA5 with pressure level has fewer layers than ERA-Interim, especially in the lower atmosphere. The boundary layer is not well described in some regions of Antarctica. Then, the vertical interpolation in the WRF model may not be very suitable over high altitude regions on Antarctica when derived from fields with fewer layers of pressure

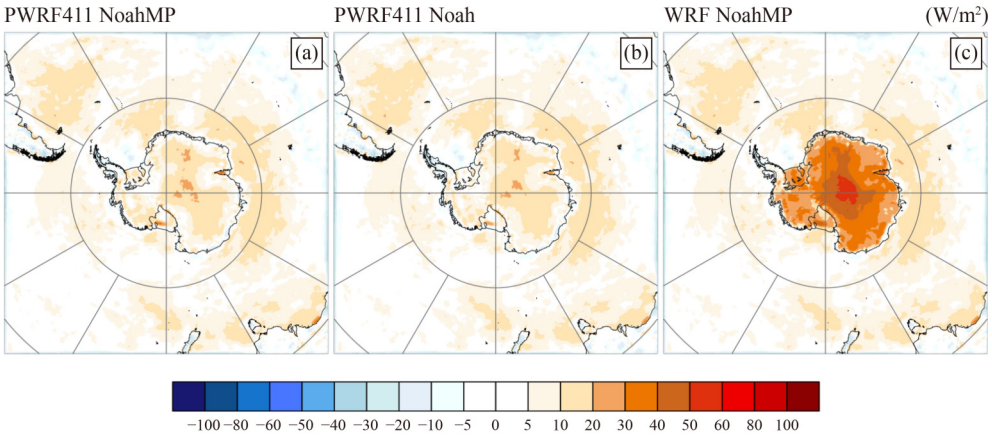

Fig. 12 The bias (Model versus ERA-Interim forecast) of longwave radiation at the surface down for (a) PWRF411 NoahMP, (b) PWRF411 Noah, and (c) WRF NoahMP in January 2009.

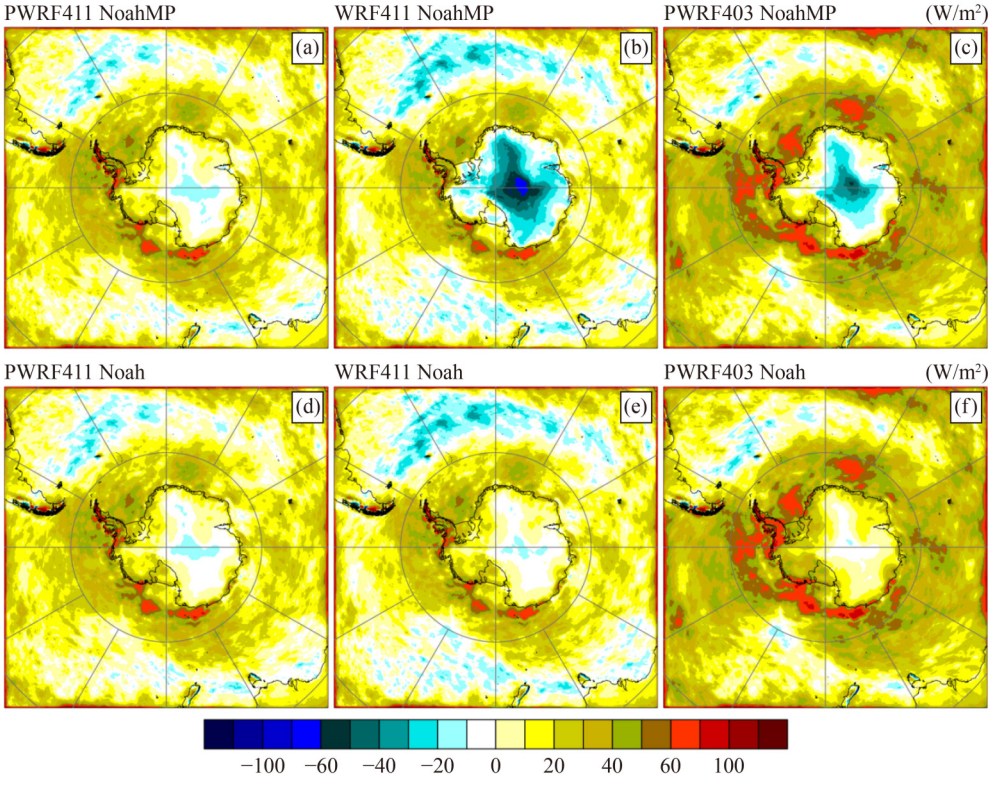

Fig. 13 The bias (Model versus ERA-Interim forecast) of shortwave radiation at the surface down for (a) PWRF411 NoahMP, (b) WRF411 NoahMP, (c) PWRF403 NoahMP, (d) PWRF411 Noah, (e) WRF411 Noah, and (f) PWRF403 Noah in January 2009.

level data. This indicates that the use of reduced vertical grid spacing for ERA5 pressure level data really penalizes the model.

4.2.3 FM and CM Evaluation

On multiple occasions, the short-term forecasting capabilities of Polar WRF were verified and showed excellent skills. (e.g., Hines and Bromwich, 2008; Bromwich et al., 2009; Hines et al., 2011; Wilson et al., 2011; Bromwich et al., 2013b). To examine the model performance of long-term simulations such as "month range" or "sub-seasonal range", PWRF411 was also tested with CM simulations. As discussed in Section 4.1, the evaluation of CM simulations between NoahMP and Noah produced similar results. Consequently, only NoahMP is shown in the evaluation for FM and CM simulations in detail (referred to as MPFM and MPCM, respectively). Here, the bias and RMSE for the 2 m temperature, 2 m dew point, surface pressure, and 10 m wind speed were calculated by month and season for analysis. As the statistical data demonstrate the same conclusions, for conciseness, only

Table 6 January and July correlations (CORR), RMSE, and bias from the Reanalysis, PWRF 48 hour forecast (spin-up 24 hour) with ERA5 (E5) and ERA-Interim (EI) (*n* = number of stations, domain-averaged statistics)

	Reanalysis		NoahMP		Noah	
	E5	**EI**	**E5**	**EI**	**E5**	**EI**
			January			
2 m Temperature (°C) *n* = 87						
CORR	0.80	0.85	0.74	0.73	0.74	0.72
BIAS	−0.97	−0.20	−1.56	−1.15	−1.17	−0.84
RMSE	2.14	2.00	2.80	2.88	2.92	3.00
2 m Dew point (°C) *n* = 70						
CORR	0.78	0.72	0.72	0.71	0.72	0.70
BIAS	−0.54	0.09	−0.21	0.26	−0.39	−0.03
RMSE	2.53	2.73	2.92	3.03	2.87	3.00
Surface pressure (hPa) *n* = 83						
CORR	0.97	0.99	0.99	0.99	0.99	0.99
BIAS	1.43	0.61	1.06	0.52	1.00	0.48
RMSE	2.21	1.70	1.94	1.65	1.91	1.65
10 m Wind speed (m/s) *n* = 80						
CORR	0.69	0.66	0.65	0.62	0.65	0.62
BIAS	−0.70	−0.58	0.30	−0.02	0.31	0.02
RMSE	2.88	2.98	2.94	3.04	2.94	3.04
			July			
2 m Temperature (°C) *n* = 67						
CORR	0.89	0.89	0.83	0.84	0.84	0.84
BIAS	0.22	−0.06	−1.25	−0.96	−1.09	−0.77
RMSE	2.98	3.21	3.94	4.03	3.90	3.99
2 m Dew point (°C) *n* = 57						
CORR	0.86	0.83	0.84	0.83	0.83	0.83
BIAS	−0.65	−0.69	−1.17	−0.78	−1.53	−0.91
RMSE	3.59	4.20	3.97	4.03	4.10	4.01
Surface pressure (hPa) *n* = 66						
CORR	0.99	0.99	0.99	0.99	0.99	0.99
BIAS	2.60	1.46	2.20	1.41	2.18	1.40
RMSE	3.36	2.57	2.98	2.43	2.97	2.42
10 m Wind speed (m/s) *n* = 59						
CORR	0.72	0.67	0.69	0.68	0.69	0.68
BIAS	−0.67	0.05	1.55	0.99	1.82	1.23
RMSE	3.98	4.15	4.51	4.34	4.62	4.41

the four seasonal results averaged every three months (JJA, SON, DJF, and MAM) are given in Table 7. Most of the year, the MPCM has the same or even higher correlation with a smaller absolute value of bias and RMSE than those of the MPFM, except for the 10 m wind speed during the SON and DJF. Statistics indicate that the model exhibits good performance in the long term forecast. Furthermore, the 2 m temperature diurnal cycle analysis for each station demonstrates that the CM simulation can usually weaken the intense temperature drop for coastal stations. In Fig. 14, the diurnal 2 m temperature cycle for Halley, Neumayer, Syowa, Mawson, Davis, Mirny, Casey, Dumont D'Urville, and McMurdo are shown. First, both NoahMP and Noah, CM (solid line) always shows a weaker diurnal cycle than FM (see dotted line). This is very likely because of the suitable nudging setting used, which can improve the performance of PWRF (Cassano et al., 2011). Based on

Table 7 Seasonal domain-averaged statistics for NoahMP of FM and CM

	MPFM				MPCM			
	JJA	**SON**	**DJF**	**MAM**	**JJA**	**SON**	**DJF**	**MAM**
2 m Temperature (°C)								
CORR	0.84	0.84	0.77	0.82	0.84	0.84	0.77	0.82
BIAS	−0.99	−0.58	−1.18	−1.21	−0.92	−0.50	−0.96	−1.02
RMSE	4.03	3.34	3.06	4.04	3.94	3.25	2.88	3.87
2 m Dew point (°C)								
CORR	0.83	0.80	0.72	0.81	0.83	0.80	0.73	0.81
BIAS	−0.84	−0.16	0.28	−1.22	−0.76	−0.08	0.51	−1.03
RMSE	4.19	3.50	3.25	4.28	4.15	3.45	3.16	4.16
Surface pressure (hPa)								
CORR	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99
BIAS	1.38	1.15	0.67	1.08	1.36	1.11	0.65	1.06
RMSE	2.28	2.00	1.76	2.14	2.27	1.99	1.75	2.13
10 m Wind speed (m/s)								
CORR	0.66	0.66	0.65	0.66	0.66	0.66	0.64	0.66
BIAS	0.87	0.41	−0.01	0.32	0.87	0.42	−0.02	0.31
RMSE	4.20	3.88	3.20	4.14	4.19	3.88	3.21	4.14

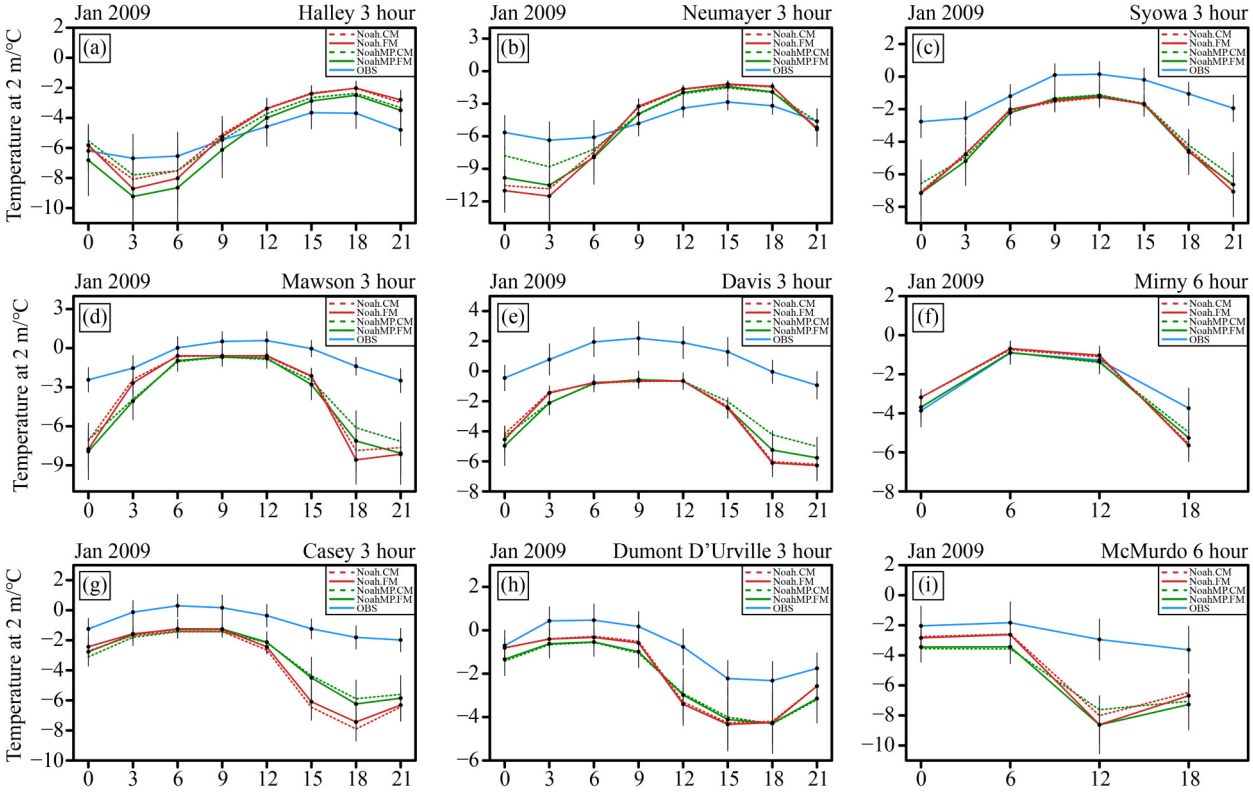

Fig. 14 Diurnal 2 m temperature cycle of nine stations located at Antarctic coastline for observation (solid blue line), PWRF411 NoahMP simulated with FM and CM (solid green line and dashed green line), and PWRF411 Noah simulated with FM and CM (solid red line and dashed red line), respectively. The standard deviation plotted as vertical error bar for observation (OBS), NoahMP and Noah in FM. Note that observations are every 3 h at Halley, Neumayer, Syowa, Mawson, Davis, Casey, and Dumont D'Urville while every 6 h at Mirny and McMurdo.

the series of test experiments, we found that nudging of long-term forecasts can effectively improve temperature forecast skills at the bottom and top levels, especially at the top of the model. The cold bias with ERA-Interim reanalysis (absence of observations) at the model top level reduced sharply over the Antarctic. Then, regardless of whether PWRF4111 was forecasted with FM or CM, NoahMP illustrates a weaker diurnal cycle than Noah, as mentioned in Section 4.1.1.1

5 Summary and conclusions

This paper has investigated the downscaling simulation performance of PWRF411 over the Antarctic and the Southern Ocean. State variables from short-term forecasts (FM), long-term forecasts (CM) during an annual cycle and different versions in two seasons have been examined, respectively. The results demonstrate that the recent version of PWRF shows good forecast skills and better than those of earlier Antarctic simulations (B13). Although NoahMP and Noah simulations reached similar conclusions, NoahMP illustrates slightly better capabilities than Noah, particularly over the ice sheet.

For the surface temperature, PWRF411 shows a cold domain-averaged 2 m temperature bias and variations in spatial analysis. Coastal regions over the Antarctic and the Southern Ocean are usually cold, but the reverse situation occurs in the Antarctic interior. Over the ice sheet, NoahMP with albedo modification removed the strong warm bias, showing the best performance in summer, which was supported by further temperature diurnal cycle and surface radiation balance analysis. Not surprisingly, the 2 m dew point shows a similar result but with more changes near summertime. For the surface pressure, PWRF411 always predicted values that were the best of the near-surface variables. Correlations typically range above ~0.99 with a slight positive bias throughout the entire year matching the cold temperature bias. For the surface wind, there is a slight indication of a strong 10 m wind speed in most of months. The maximum bias and RMSE appearing in July are likely because wind speeds reach a peak during this time. There is a deficit in the downwelling longwave at the coastline but an excess in other areas. The downwelling shortwave is excessive in the South Pole, the West Antarctic, and coastline but is insufficient on the East Antarctic Plateau during summertime. These characteristics for LWD and SWD are also consistent with previous findings (Bromwich et al., 2013b). Additionally, upper air analysis illustrates that PWRF411 accurately describes the characteristics of the atmospheric vertical profiles. The vertical profiles of temperature and horizontal wind speed clearly show the tropopause (~300 hPa), especially during summer. Although statistics show progress, relative humidity still faces challenges both in the model and observations in the Antarctic.

Moreover, this study also extends the findings of model performance by sensitivity experiments for different model versions, driving data and simulation modes.

First, the comparison of different versions shows that the new version of PWRF411 with modified NoahMP performed better than other versions.

Second, simulations driven by ERA5 (37 pressure vertical levels, ~31 km for horizontal resolution) did not show better performance than ERA-Interim (60 vertical levels, ~79 km for horizontal resolution) in our study. This suggests that the use of reduced vertical grid spacing in the driving data really penalizes the model.

Third, long-term forecasting performed well. The assessment of model long-term simulations proved that an appropriate nudging setting plays an important role in promoting model performance. It keeps the simulations close to the forcing data when the lateral boundary is updated. This means that regional forecast results are sensitive to the global model capability (lateral boundary) for long-term forecasts.

To summarize, notwithstanding the unavoidable ambiguities, these characteristics examined in this study provide a benchmark to improve the model and guidance for further application of Polar WRF in the Antarctic. The study also extends the findings of model shortcomings, such as the strong diurnal cycle in coastal areas compared to the observations. Further model development is needed to solve this problem. Moreover, data assimilation also will be developed.

Acknowledgments This research was supported by the Chinese Academy of Sciences (No. XDA20060501) and the National Natural Science Foundation of China (Grant No. 91937000) to the first two authors. The other co-authors were supported by the Office of Naval Research (ONR) (No. N00014-18-1-2361). We are grateful to the Antarctic Meteorological Research Center (AMRC) at the University of Wisconsin-Madison and the OGIMET database for the surface observations. We also would like to thank University of Wyoming and the Antarctic Meteo-Climatological Observatory funded by the Italian National Program of Antarctic Research for the upper-air observation data. We appreciate the use of BSRN stations data for downwelling surface longwave and shortwave radiation. Thanks for the comments from the reviewers and editors that substantially improved this article. Contribution No. 1616 of Byrd Polar and Climate Research Center.

References

Alley R B, Emanuel K A, Zhang F (2019). Advances in weather prediction. Science, 363(6425): 342–344

Bauer P, Thorpe A, Brunet G (2015). The quiet revolution of numerical weather prediction. Nature, 525(7567): 47–55

Berrisford P, Dee D P, Poli P, Brugge R, Fielding M, Fuentes M, Kållberg P W, Kobayashi S, Uppala S, Simmons A (2011). The ERA-Interim archive Version 2.0. ECMWF

Bracegirdle T J, Marshall G J (2012). The reliability of antarctic tropospheric pressure and temperature in the latest global reanalyses. J Clim, 25(20): 7138–7146

Bromwich D H, Bai L, Bjarnason G G (2005). High-resolution regional climate simulations over iceland using Polar MM5*. Mon Weather

Rev, 133(12): 3527–3547

Bromwich D H, Hines K M, Bai L (2009). Development and testing of polar weather research and forecasting model: 2. Arctic Ocean. J Geophys Res, 114(D8): D08122

Bromwich D H, Nicolas J P, Hines K M, Kay J E, Key E L, Lazzara M A, Lubin D, McFarquhar G M, Gorodetskaya I V, Grosvenor D P, Lachlan-Cope T, van Lipzig N P M (2012). Tropospheric clouds in Antarctica. Rev Geophys, 50(1): RG1004

Bromwich D H, Nicolas J P, Monaghan A J, Lazzara M A, Keller L M, Weidner G A, Wilson A B (2013a). Central West Antarctica among the most rapidly warming regions on Earth. Nat Geosci, 6(2): 139–145

Bromwich D H, Nicolas J P, Monaghan A J, Lazzara M A, Keller L M, Weidner G A, Wilson A B (2014). Erratum: corrigendum: Central West Antarctica among the most rapidly warming regions on Earth. Nat Geosci, 7(1): 76

Bromwich D H, Otieno F O, Hines K M, Manning K W, Shilo E (2013b). Comprehensive evaluation of polar weather research and forecasting model performance in the Antarctic J Geophys Res D Atmospheres, 118(2): 274−292

Bromwich D H, Wilson A B, Bai L, Liu Z, Barlage M, Shih C F, Maldonado S, Hines K M, Wang S H, Woollen J, Kuo B, Lin H C, Wee T K, Serreze M C, Walsh J E (2018). The arctic system reanalysis, Version 2. Bull Am Meteorol Soc, 99(4): 805–828

Cassano J J, Higgins M E, Seefeldt M W (2011). Performance of the weather research and forecasting model for month-long Pan-Arctic simulations. Mon Weather Rev, 139(11): 3469–3488

Chen S Y, Wee T K, Kuo Y H, Bromwich D H (2014). An impact assessment of GPS radio occultation data on prediction of a rapidly developing cyclone over the Southern Ocean. Mon Weather Rev, 142(11): 4187–4206

DeConto R M, Pollard D (2016). Contribution of Antarctica to past and future sea-level rise. Nature, 531(7596): 591–597

Glisan J M, Gutowski W J Jr, Cassano J J, Higgins M E (2013). Effects of spectral nudging in WRF on Arctic temperature and precipitation simulations. J Clim, 26(12): 3985–3999

Guo Z, Bromwich D H, Cassano J J (2003). Evaluation of Polar MM5 simulations of antarctic atmospheric circulation. Mon Weather Rev, 131(2): 384–411

Hines K M, Bromwich D H (2008). Development and testing of Polar Weather Research and Forecasting (WRF) Model. Part I: Greenland ice sheet meteorology. Mon Weather Rev, 136(6): 1971–1989

Hines K M, Bromwich D H (2017). Simulation of late summer Arctic clouds during ASCOS with Polar WRF. Mon Weather Rev, 145(2): 521–541

Hines K M, Bromwich D H, Bai L, Bitz C M, Powers J G, Manning K W (2015). Sea ice enhancements to Polar WRF. Mon Weather Rev, 143(6): 2363–2385

Hines K M, Bromwich D H, Bai L S, Barlage M, Slater A G (2011). Development and testing of Polar WRF. Part III: Arctic land. J Clim, 24(1): 26–48

Hines K M, Bromwich D H, Wang S H, Silber I, Verlinde J, Lubin D (2019). Microphysics of summer clouds in central west Antarctica simulated by Polar WRF and AMPS. Atmos Chem Phys Discuss,

Iacono M J, Delamere J S, Mlawer E J, Shephard M W, Clough S A, Collins W D (2008). Radiative forcing by long-lived greenhouse gases: calculations with the AER radiative transfer models. J Geophys Res, 113(D13): D13103

Jones P D, Lister D H (2015). Antarctic near-surface air temperatures compared with ERA-Interim values since 1979. Int J Climatol, 35(7): 1354–1366

Jung T, Bauer P, Goessling H, Gordon N, Klebe S, Bromwich D, Doblas-Reyes F, Day J, Fairall C, Holland M, Iversen T, Lemke P, Mills B, Nurmi P, Perovich D, Reid P, Renfrew I, Smith G, Svensson G, Tolstykh M. (2015). The WWRP Polar Prediction Project (PPP). In: Seamless Prediction of The Earth System: From Minutes to Months, WMO-No. 1156, Geneva, WMO

Kain J S (2004). The Kain-Fritsch convective parameterization: an update. J Appl Meteorol, 43(1): 170–181

Kennicutt M C 2nd, Chown S L, Cassano J J, Liggett D, Massom R, Peck L S, Rintoul S R, Storey J W V, Vaughan D G, Wilson T J, Sutherland W J (2014). Polar research: six priorities for Antarctic science. Nature, 512(7512): 23–25

Kennicutt M C II, Chown S L, Cassano J J, Liggett D, Peck L S, Massom R, Rintoul S R, Storey J, Vaughan D G, Wilson T J, Allison I, Ayton J, Badhe R, Baeseman J, Barrett P J, Bell R E, Bertler N, Bo S, Brandt A, Bromwich D, Cary S C, Clark M S, Convey P, Costa E S, Cowan D, Deconto R, Dunbar R, Elfring C, Escutia C, Francis J, Fricker H A, Fukuchi M, Gilbert N, Gutt J, Havermans C, Hik D, Hosie G, Jones C, Kim Y D, Le Maho Y, Lee S H, Leppe M, Leitchenkov G, Li X, Lipenkov V, Lochte K, López-Martínez J, Lüdecke C, Lyons W, Marenssi S, Miller H, Morozova P, Naish T, Nayak S, Ravindra R, Retamales J, Ricci C A, Rogan-Finnemore M, Ropert-Coudert Y, Samah A A, Sanson L, Scambos T, Schloss I R, Shiraishi K, Siegert M J, Simões J C, Storey B, Sparrow M D, Wall D H, Walsh J C, Wilson G, Winther J G, Xavier J C, Yang H, Sutherland W J (2015). A roadmap for Antarctic and Southern Ocean science for the next two decades and beyond. Antarct Sci, 27(1): 3–18

Kennicutt M C II, Kim Y D, Rogan-Finnemore M, Anandakrishnan S, Chown S L, Colwell S, Cowan D, Escutia C, Frenot Y, Hall J, Liggett D, Mcdonald A J, Nixdorf U, Siegert M J, Storey J, Wåhlin A, Weatherwax A, Wilson G S, Wilson T, Wooding R, Ackley S, Biebow N, Blankenship D, Bo S, Baeseman J, Cárdenas C A, Cassano J, Danhong C, Dañobeitia J, Francis J, Guldahl J, Hashida G, Corbalán L J, Klepikov A, Lee J, Leppe M, Lijun F, López-Martinez J, Memolli M, Motoyoshi Y, Bueno R M, Negrete J, Cárdenes M A O, Silva M P, Ramos-Garcia S, Sala H, Shin H, Shijie X, Shiraishi K, Stockings T, Trotter S, Vaughan D G, De Menezes J V D U, Vlasich V, Weijia Q, Winther J G, Miller H, Rintoul S, Yang H (2016). Delivering 21st century Antarctic and Southern Ocean science. Antarct Sci, 28(6): 407–423

Mayewski P A, Meredith M P, Summerhayes C P, Turner J, Worby A, Barrett P J, Casassa G, Bertler N A N, Bracegirdle T, Naveira Garabato A C, Bromwich D, Campbell H, Hamilton G S, Lyons W B, Maasch K A, Aoki S, Xiao C, van Ommen T (2009). State of the Antarctic and Southern Ocean climate system. Rev Geophys, 47(1): RG1003

Miguez-Macho G, Stenchikov G L, Robock A (2004). Spectral nudging to eliminate the effects of domain position and geometry in regional climate model simulations. J Geophys Res D Atmosph, 109: D13104

Monaghan A, Bromwich D (2008). Global warming at the poles. Nat Geosci, 1(11): 728–729

Nakanishi M, Niino H (2009). Development of an improved turbulence closure model for the atmospheric boundary layer. J Meteorol Soc Jpn, 87(5): 895–912

Niu G Y, Yang Z L, Mitchell K E, Chen F, Ek M B, Barlage M, Kumar A, Manning K, Niyogi D, Rosero E, Tewari M, Xia Y (2011). The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements. J Geophys Res D Atmospheres, 116(D12): D12109

Olson J B, Kenyon J S, Angevine W A, Brown J M, Pagowski M, Sušelj K (2019). A description of the MYNN-EDMF scheme and the coupling to other components in WRF–ARW

Powers J G, Klemp J B, Skamarock W C, Davis C A, Dudhia J, Gill D O, Coen J L, Gochis D J, Ahmadov R, Peckham S E, Grell G A, Michalakes J, Trahan S, Benjamin S G, Alexander C R, Dimego G J, Wang W, Schwartz C S, Romine G S, Liu Z, Snyder C, Chen F, Barlage M J, Yu W, Duda M G (2017). The weather research and forecasting model: overview, system efforts, and future directions. Bull Am Meteorol Soc, 98(8): 1717–1737

Powers J G, Manning K W, Bromwich D H, Cassano J J, Cayette A M (2012). A decade of Antarctic science support through AMPS. Bull Am Meteorol Soc, 93(11): 1699–1712

Scambos T A, Bell R E, Alley R B, Anandakrishnan S, Bromwich D H, Brunt K, Christianson K, Creyts T, Das S B, DeConto R, Dutrieux P, Fricker H A, Holland D, MacGregor J, Medley B, Nicolas J P, Pollard D, Siegfried M R, Smith A M, Steig E J, Trusel L D, Vaughan D G, Yager P L (2017). How much, how fast? A science review and outlook for research on the instability of Antarctica's Thwaites Glacier in the 21st century Global Planet Change, 153: 16–34

Scott R C, Nicolas J P, Bromwich D H, Norris J R, Lubin D (2019). Meteorological drivers and large-scale climate forcing of West Antarctic surface melt. J Clim, 32(3): 665–684

Urban M C, Bocedi G, Hendry A P, Mihoub J B, Pe'er G, Singer A, Bridle J R, Crozier L G, De Meester L, Godsoe W, Gonzalez A, Hellmann J J, Holt R D, Huth A, Johst K, Krug C B, Leadley P W, Palmer S C F, Pantel J H, Schmitz A, Zollner P A, Travis J M (2016). Improving the forecast for biodiversity under climate change. Science, 353(6304): aad8466

Wilson A B, Bromwich D H, Hines K M (2011). Evaluation of Polar WRF forecasts on the Arctic system reanalysis domain: surface and upper air analysis. J Geophys Res, 116(D11): D11112

Wilson A B, Bromwich D H, Hines K M (2012). Evaluation of Polar WRF forecasts on the Arctic System Reanalysis Domain: 2. atmospheric hydrologic cycle. J Geophys Res D Atmosph, 117: D04107

Xue J, Bromwich D H, Xiao Z, Bai L (2021). Impacts of initial conditions and model configuration on simulations of polar lows near Svalbard using Polar WRF with 3DVAR. Q J R Meteorol Soc, 147(740): 3806–3834

Zou X, Bromwich D H, Nicolas J P, Montenegro A, Wang S H (2019). West Antarctic surface melt event of January 2016 facilitated by föhn warming. Q J R Meteorol Soc, 145(719): 687–704

AUTHOR BIOGRAPHIES

Jianjun Xue, Ph.D. candidate of State Key Laboratory of Numerical Modeling for Atmospheric Sciences and Geophysical Fluid Dynamics, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China. He mainly focuses on polar meteorology and climatology being studied using climate models. He was working for the polar version of the regional WRF model that is being developed, tested, and applied as a visiting scholar at The Ohio State University, USA from August 2018 to July 2020. His e-mail is jianjxue@hotmail.com.

Ziniu Xiao, received his Ph.D. Degree in Atmospheric Science from Institute of Atmospheric Physics, Chinese Academy of Sciences. Beijing, China, in 2006. He is the chief scientist of National Basic Research Program of China Program. Professor and director of State Key Laboratory of Numerical Modeling for Atmospheric Sciences and Geophysical Fluid Dynamics (LASG), Institute of Atmospheric Physics. He mainly focuses on climate dynamics, climate prediction and weather forecast, solar impact on climate systems. Dr. Xiao's e-mail is xiaozn@lasg.iap.ac.cn.

David H. Bromwich, received his Ph.D. Degree in meteorology from the University of Wisconsin-Madison, USA, in 1979. Research Professor, Senior Research Scientist, Atmospheric Sciences Program, Dept. of Geography, Byrd Polar and Climate Research Center of The Ohio State University. He focuses on global climate change in high latitudes resulting from local and tropical influences and uses climate models and atmospheric reanalyses as well as the polar version of the regional WRF model that is being developed, tested, and applied to climate variability and change problems in both polar regions. Dr. Bromwich's e-mail is bromwich.1@osu.edu.

Lesheng Bai, Senior Research Associate of Polar Meteorology Group, Byrd Polar and Climate Research Center of The Ohio State University. He focuses on the polar version of the regional WRF model that is being developed, tested, and applied to climate variability and change problems in both polar regions. His e-mail is bailesheng@hotmail.com.