Antarctic Verification of the Australian Numerical Weather Prediction Model

BENJAMIN J. E. SCHROETER

Institute for Marine and Antarctic Studies, Hobart, Tasmania, Australia

PHIL REID

Bureau of Meteorology, Hobart, Tasmania, Australia

NATHANIEL L. BINDOFF

Antarctic Climate and Ecosystem Cooperative Research Centre, Hobart, Tasmania, Australia

KELVIN MICHAEL

University of Tasmania, Hobart, Tasmania, Australia

ABSTRACT

The Australian Community Climate and Earth-System Simulator-Global (ACCESS-G) features an atmosphere-only numerical weather prediction (NWP) suite used operationally by the Australian Bureau of Meteorology to forecast weather conditions for the Antarctic. The current operational version of the forecast model, the Australian Parallel Suite v2 (APS2), has been used operationally since early 2016. To date, the performance of the model has been largely unverified for the Antarctic and anecdotal reports suggest challenges for model performance in the region. This study investigates the performance of ACCESS-G south of 50°S over 2017 and finds that model performance degrades toward the poles and in proportion to forecast horizon against a range of performance metrics. The model exhibits persistent negative surface and mean sea level pressure biases around the Adelie Land coast, which is linked to the underrepresentation of model winds to the west, and driven by positive screen temperature biases that inhibit modeled katabatic outflow. These results suggest that an improved representation of boundary layer parameterization could be implemented to improve model performance in the region.

1. Introduction

Research and logistical activities on and around the Antarctic continent are critically dependent on the provision of reliable weather forecasts. Base operations, including research expeditions, are highly dependent on forecast guidance, as hazardous conditions can develop quickly, limiting accessibility to field sites and potentially threatening the safety of researchers (Powers et al. 2003). Aviation and shipping endeavors require forecasting services for situational awareness hours or even days in advance, and decisions made as a result of these forecasts have substantial financial and safety implications. The capacity of national meteorological agencies to issue timely, accurate weather forecast guidance for the Antarctic region is therefore of paramount importance.

However, observational records of the weather and climate of Antarctica and the Southern Ocean are temporally limited and spatially sparse. Since the 1950s, automatic weather stations (AWSs) have been installed on the continent to take surface meteorological measurements, and in recent decades to also take surface chemistry measurements (Lazzara et al. 2012). Continuing to install new AWSs and maintain existing sites requires considerable financial and logistical investment. Furthermore, not all AWSs are recognized or used by every national Antarctic program, and they produce data of variable quality. Logistical and financial constraints on the number of stations that can be serviced per season result in a substantial proportion of sites being left for

---

ᵃ Denotes content that is immediately available upon publication as open access.

Corresponding author: Benjamin J. E. Schroeter, benjamin.schroeter@utas.edu.au

several years between maintenance visits. During periods of this length, the accumulation of snow changes the elevation of the instruments above the surface, and some stations can even be entirely buried beneath snow.

AWSs are sometimes installed on surfaces other than rock, such as on ice shelves, which move over time and require additional effort to ensure measurements of elevation and location are adequately maintained (Lazzara et al. 2012). The AWS network that currently exists is generally concentrated in regions near manned stations, such as around the West Antarctic Peninsula, the Ross Ice Shelf, and in parts of the East Antarctic coastline such as Adelie Land (Fig. 1). However, large regions of the continental interior and coastline, as well as the sea ice and surrounding ocean, remain without any AWS observational records. While satellite observations have provided enhanced spatiotemporal observational coverage of the region, the utility of these observations is limited, such as by cloud or temporal coverage (Comiso 2000; Walton 2013).

The lack of a long-term, continuous observational record of key weather variables places constraints upon weather forecasting and research, which necessitates targeted and internationally collaborative observing campaigns. One such campaign is the Year of Polar Prediction (YOPP) Special Observing Period (SOP), which ran from 16 November 2018 to 15 February 2019 (see Goessling et al. 2016) and aims to populate the observational record with enhanced observations over an extended period of time.

While the output of NWP can be used to fill gaps in the observational record, a higher level of caution must be taken with NWP output than in midlatitude regions, since the sparsity of observational data can lead to model drift (Connolley and Harangozo 2001) or a greater influence of the model background (i.e., a prior forecast). The spatiotemporal variability of NWP output has been shown to have differing levels of predictive skill between the mid- and high latitudes, as well as varying both horizontally and vertically throughout the atmosphere (Bengtsson 1991). The predictive skill of NWP is sensitive to data paucity, whereby fewer observations can lead to a greater contribution of the model prior during assimilation, and greater reliance on the model itself during verification. Heat flux and momentum energy transfers, and errors in the model parameterization of physical processes such as cloud microphysics also have substantial impacts on the skill of NWP (Bauer et al. 2015). The verification of NWP model output is therefore more challenging in regions of lower observational coverage, such as in the high southern latitudes.

Traditional NWP output verification uses a range of metrics and skill scores (Wilks 2011; Bauer et al. 2015; Jung and Matsueda 2016) for a range of meteorological parameters such as mean sea level pressure (MSLP), geopotential height at 500 hPa, and surface winds and temperatures (WMO 2015). MSLP is typically used for the identification of high or low pressure systems, which are essential for the forecasting of both the type and severity of weather phenomena; it is also less dominated by biases in orography (Bracegirdle and Marshall 2012) and provides insight into atmospheric conditions both at the surface and throughout the atmospheric column above. The standard geopotential height for analysis in the Antarctic region is 500 hPa as it is the first mandatory reporting geopotential height level that is located everywhere above the ice surface (Pendlebury et al. 2003). In addition, as the 500-hPa surface is also above the planetary boundary layer in the free atmosphere, its flow is in near-geostrophic balance and not influenced by surface effects such as friction and shearing stresses (Kaimal and Finnigan 1994). The 500-hPa surface is used as a general performance indicator independent of boundary layer and surface parameterizations, which become less important at these heights for variables such as temperature (Bracegirdle and Marshall 2012).

Real-time forecasting for the Antarctic relies heavily on both global and limited-area (or mesoscale) atmospheric models (LAMs), such as the popular Antarctic Mesoscale Prediction System (AMPS). AMPS was first implemented in late 2000 to provide experimental real-time meteorological forecasts for the Antarctic region (Powers et al. 2003) and has been run operationally as a real-time implementation of the Weather Research and Forecasting (WRF) Model (Skamarock et al. 2008) since 2008. Initial conditions in AMPS are generated from the National Centers for Environmental Prediction (NCEP) global forecasting system as well as space-borne and surface observations. As a limited-area model, AMPS covers six domains: one domain that encompasses most of the Southern Ocean, another over the Antarctic continent, and four others that focus specifically on regions of interest such as the Ross Sea and South Pole (Bromwich et al. 2005). The performance of AMPS forecasts has been found to be generally strong, due in part to polar-specific modifications to the original model, including changes to the radiation scheme, incorporating fractional sea ice coverage, updates to the thermal properties of sea ice, ice and snow, and careful treatment of the Antarctic coastal topography and land surface (Powers et al. 2003).

In conjunction with LAMS, forecasting centers also provide weather forecasts for the polar regions using outputs from global prediction systems. One such model is the global variant of the Australian Community Climate and Earth-System Simulator (ACCESS); operated and
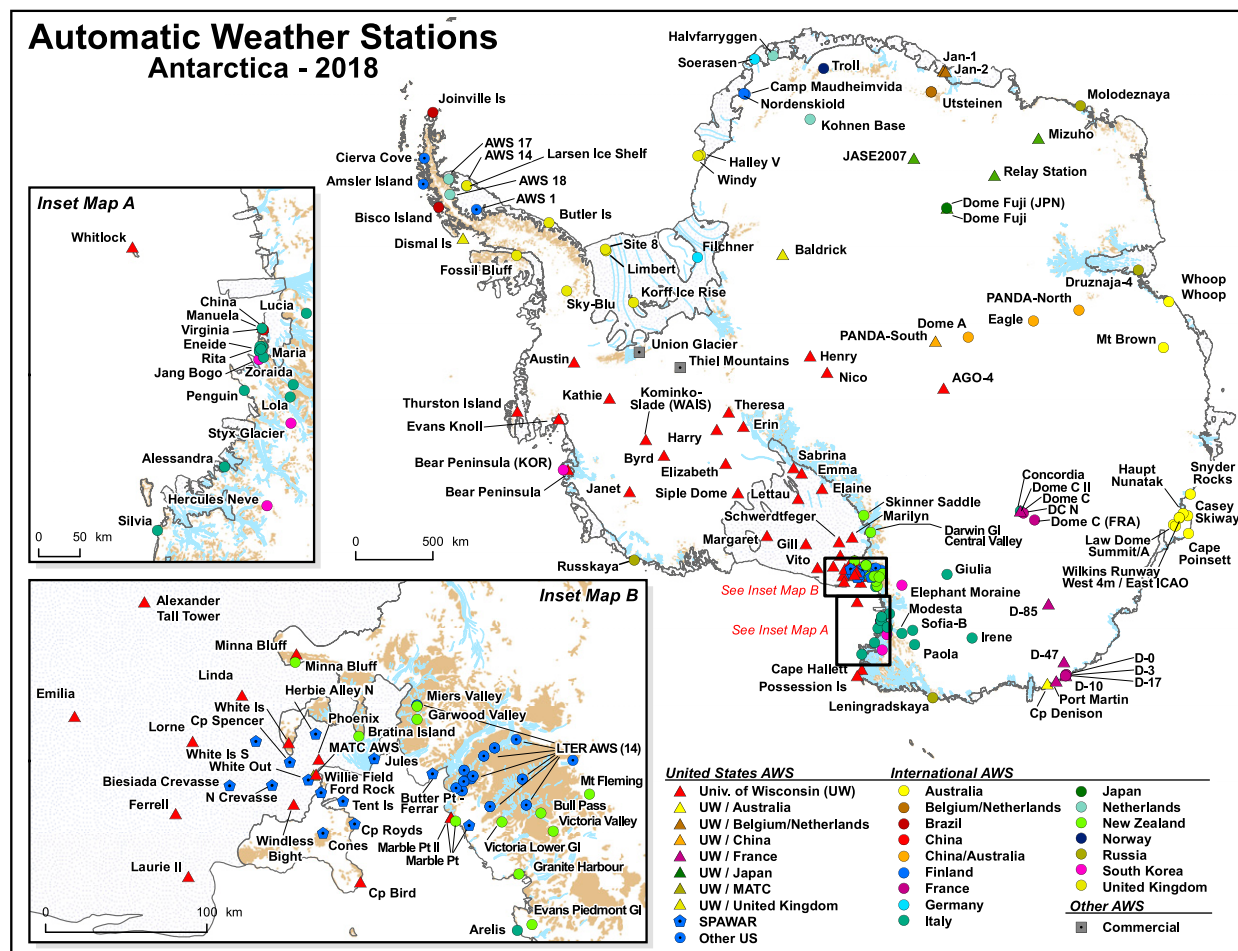
FIG. 1. Overview of Automatic Weather Stations in the Antarctic (AMRC 2018). Note that some AWSs [i.e., Mawson (Australia)] do not feature on this diagram.

maintained by the Australian Bureau of Meteorology (Puri et al. 2013). ACCESS is an atmosphere-only NWP suite built upon the Met Office (UKMO) Unified Model (UM; see Cullen 1993); using a combination of UKMO and custom components developed specifically for the Southern Hemisphere. Initial conditions are generated via a four-dimensional variational assimilation system (4DVAR; see Rawlins et al. 2007); combining quality-controlled observations with the model prior and background error covariances (Puri et al. 2013). Table 1 briefly describes general details of the model and the reader is referred to Puri et al. (2013), Australian Bureau of Meteorology (2016) and Davies et al. (2005) for more specific details regarding model implementation and physical parameterizations. ACCESS has several variants spanning global (ACCESS-G), regional (ACCESS-R), and city (ACCESS-C) domains, as well as a relocatable tropical cyclone domain (ACCESS-TC). However, as there is no polar-specific version of

ACCESS, Antarctic forecasters rely on forecast guidance from the global variant, ACCESS-G; upon which this study is focused.

The performance of ACCESS is released to the public in the form of quarterly performance statements (Wu 2015, 2016). These statements assess model skill of MSLP and 500-hPa geopotential height for the Australian verification domain for both the global model and the higher-resolution regional forecast model. In addition, these statements chart the performance of ACCESS compared to international models from other forecasting centers. While these statements focus primarily on the performance over the Australian verification domain, model forecast data remain available for the polar regions. This presents both the opportunity and the data required to assess model performance in the Antarctic.

Anecdotal reports suggest that atmosphere-only weather forecast models have unreliable performance over the

TABLE 1. ACCESS-G APS2 configuration overview. See Puri et al. (2013) and Australian Bureau of Meteorology (2016) for further details.

Atmospheric core	UKMO Unified Model (UM) version 8.2
Boundary layer	Joint U.K. Land Environment Simulator (JULES; Best et al. 2011)
Clouds	Prognostic cloud and condensate (PC2) scheme (Wilson and Ballard 1999) with cloud area correction (Boutle and Morcrette 2010)
Convection	Mass flux scheme based on Gregory and Rowntree (1990) version 4a
Data assimilation (DA)	Four-dimensional variational assimilation (4DVAR) version v29.1
DA frequency	6-hourly at 0000, 0600, 1200, 1800 UTC
Horizontal resolution	769×1024 (approx. $0.23° \times 0.35°$, 25 km)
Precipitation	Single-moment bulk microphysics scheme (Wilson and Ballard 1999) with modifications from Abel and Shipway (2007)
Radiation	Edwards and Slingo (1996) with additions outlined in Australian Bureau of Meteorology (2016)
Sea surface temperature	GAMSSA Daily Global 0.25° SST analysis (Beggs 2008)
Sea ice	NCEP 1/12° sea ice analysis
Soil moisture analysis	SURF nudges soil moisture field via screen-level analysis; once every 6 h
Vertical resolution	70 levels, model top approx. 80 km/0.009 hPa

Antarctic continent and surrounding ocean, leading to the interpretation of ACCESS-G in concert with models from other centers to confirm or repudiate the model output. It is through this approach that forecasters leverage the strengths/weaknesses of each model to find an agreement that best informs the forecasting process.

Until recently, the performance of NWP has had particular emphasis on the tropics and midlatitudes (Jung and Matsueda 2016), where societal implications are understandably weighted by the proportion of the population residing within these latitude ranges. Nevertheless, the performance of NWP toward the poles is experiencing renewed interest from the international community and National Arctic/Antarctic Programs through the World Meteorological Organization's Polar Prediction Project (PPP) and YOPP 2017–19. Furthermore, there is a particular emphasis on the verification of the complex polar environment (Casati et al. 2017) to which this study aims to contribute.

In this study, we seek to understand the degradation of performance of ACCESS-G south of 50°S through interpretation of standard evaluation techniques used operationally by the Australian Bureau of Meteorology and other forecasting centers. We focus on the S1 skill score (Teweles and Wobus 1954), mean error (ME), mean absolute error (MAE), and root-mean-square error (RMSE) to assess the accuracy of forecasts for mean sea level and surface pressure, surface (10 m) winds, screen (2 m) temperature, and geopotential height at 500 hPa. Due to the limitations of using model MSLP over the Antarctic landmass (such as vertical extrapolation from the first terrain-following model level), this study also examines surface pressure variables. Through better understanding of model performance in the region, we quantify the performance of the model over the Southern Ocean and Antarctic. In doing so, we identify notable systemic model

biases, the physical processes that drive them, and whether these biases are regionally or diurnally influenced. A better understanding of model deficiencies will potentially lead to an improved future representation of Antarctic physics and parameterizations of processes not yet fully resolved in the ACCESS-G NWP model.

2. Method

a. Data

This study uses the forecast output of MSLP, surface pressure, surface winds, screen temperature, and geopotential height and air temperature at both 500 hPa and throughout the vertical column from the operational second version of the ACCESS-G Australian Parallel Suite (APS2) between January and December 2017. Forecasts and verifying analyses were temporally aligned by first subtracting the desired forecast length (horizon) in hours from the analysis time to select the forecast file preceding the reference analysis. Then, within the forecast file, the appropriate forecast horizon was selected to align with the verifying analysis. This enabled direct comparison between an analysis and the forecast (generated prior) for the same point in time. The analyses studied were the model runtimes (0000, 0600, 1200, and 1800 UTC) as compared with the 12-, 24-, 36-, and 48-h forecast horizons. These times were selected for their applicability in short-term forecasting, potential diurnal sensitivities and computational convenience. Topographical data were taken as the model's own land elevation field, which was converted from meters to (approximate) geopotential height for plotting against model outputs via the inverse of the equations provided by NOAA (2018).

While the use of model analysis as a reference verification dataset is not ideal (as data-sparse regions effectively

leave the model verifying against itself), this approach is routine in operational forecasting (i.e., Eerola 2013) and is recommended by the WMO (2015) to standardize verification between centers for model intercomparison. Point-based observations, while preferable, undersample the forecast space (Ebert et al. 2013) and are limited to the spatial distribution of installations, which in the Antarctic are perhaps fewer and sparser than anywhere else in the world. While there are efforts to verify NWP against satellites and vice versa to achieve greater observational coverage for various meteorological parameters (Crocker and Mittermaier 2013), the readings from satellite sensors are subject to their own inherent assumptions, biases and limitations, such as the delineation of cloud cover over ice and the temperature at the surface against which a model could be verified.

b. Verification metrics

Following the guidelines set forth by the WMO (2015), this study first investigates the performance of the model by the metric of the S1 skill score (Teweles and Wobus 1954) as well as the additional metrics of mean error, mean absolute error, and root-mean-squared error. Metrics are assessed spatially over the study domain as well as as meridional/zonal averages as appropriate. The inclusion of these additional metrics allows us to assess model performance from different perspectives, while acknowledging the limitations of each metric. All of the metrics described below were computed cell-wise through the model time series, with the exception of the S1 skill score.

1) S1 SKILL SCORE

The S1 score (Teweles and Wobus 1954) determines the difference between adjoining grid cells, rather than comparing absolute values of grid cells, and while it does not reflect bias (since it does not compare absolute values), it provides insight into long-term trends in forecast accuracy (Wilks 2011). S1 skill considers the gradients of values across a two-dimensional grid in the x and y dimensions for both forecast and analysis for a particular forecast parameter (typically pressure at a given level). This contrasts with differences at specific grid coordinates between forecast and analysis used in other metrics. These gradients are then compared relative to the maximum gradients of the forecast or analysis (Jolliffe and Stephenson 2012). The S1 skill score [Eq. (1)] is sensitive to both domain size and resolution, and can be manually manipulated by a forecaster by choosing a value greater than the median of the probability distribution (Thompson and Carter 1972; Jolliffe and Stephenson 2012). However, in this study we use raw model analyses and forecasts and hence there is no manual manipulation:

$$S1 = 100 \frac{\sum w_i (e_g)_i}{\sum w_i (G_l)_i},\qquad(1)$$

with

$$w_i = \cos(\phi_i),\qquad(2)$$

$$e_g = \left|\frac{\delta(f-o)}{\delta x}\right| + \left|\frac{\delta(f-o)}{\delta y}\right|,\quad \text{and}\qquad(3)$$

$$G_l = \max\left(\left|\frac{\delta f}{\delta x}\right|,\left|\frac{\delta o}{\delta x}\right|\right) + \max\left(\left|\frac{\delta f}{\delta y}\right|,\left|\frac{\delta o}{\delta y}\right|\right).\qquad(4)$$

Here w_i is applied as a weight term to reduce meridional convergence and to normalize cells across latitude ranges.

The term ϕ_i is the latitude of the current cell, f is the forecast value for the parameter of interest, o is the observed (or reference analysis) value for the parameter of interest, e_g is the (forecast − observed) pressure difference gradient, and G_l is the larger of either the forecast gradient or the observed gradient (Thompson and Carter 1972). Both e_g and G_l are solved first by computing difference fields, then the equation is solved using adjacent pairs in both the x and y dimensions to yield the S1 skill. S1 is in the range [0, ∞], with 0 indicating a perfect score. The observed value used in the calculation is typically taken as the analysis for spatial, grid-based verification and a physical observation in point-based verification (WMO, 2015).

In this study, we calculated the S1 skill of the model over the Antarctic (50°–90°S) and global domains to compare model performance and confirm that our methods were comparable with those used operationally at the bureau. Further to this, we calculated the S1 skill of the model over each global latitude band (domains of the full longitudinal range for each latitude) to observe the meridional performance of the model and any idiosyncrasies of a metric reliant on grid structure. We present the results of the meridional S1 performance in this study.

The S1 skill score possesses some undesirable qualities described by Wilks (2011), such as the lack of importance placed on the magnitude of forecast pressures, the lack of bias reflected in the metric, seasonality of performance (where summer scores tend to be worse and therefore challenging to interpret for annual time series), and sensitivity to domain size and grid structure. While considered by some to be a legacy metric that has fallen out of favor resulting from these qualities (Wilks 2011), S1 is still used operationally within the Bureau to continue the historical time series of model improvement over time. Cosine latitude weighting is not

used operationally by the Bureau over the Australian domain; however, we have used it in this study to adhere as closely as possible to the WMO specifications.

2) MEAN ERROR

The mean error, defined as

$$ME = \frac{1}{n}\sum_{i=1}^{n}(f_i - o_i), \tag{5}$$

was used to observe the average error through time at each grid point. This metric was included to establish a more general assessment of model performance, however, it is subject to issues of centrality from extreme positive or negative values that may shift the mean of the metric toward zero.

3) MEAN ABSOLUTE ERROR

The mean absolute error, defined as

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|f_i - o_i|, \tag{6}$$

is a computationally inexpensive metric that is resistant to issues of centrality and less sensitive to outliers than second-order statistics such as mean-squared error (Jolliffe and Stephenson 2012). MAE is useful when the magnitude, but not the sign of the error, is important as it does not discriminate between positive and negative errors. We have included it to observe error magnitudes, however, higher-order statistics such as mean-squared error are more commonplace in operational verification.

4) ROOT-MEAN-SQUARED ERROR

The root-mean-squared error, defined as

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(f_i - o_i)^2}, \tag{7}$$

was selected over the commonly used mean-squared error metric as it retains the units of the original data. RMSE does not suffer from centrality issues from extreme positive or negative values. Instead, RMSE is sensitive to these outliers and heavily penalizes extreme errors rather than diluting them through the calculation of the mean. We have used this statistic to detect local error maxima that may be the result of regional or diurnal influence.

This metric is most appropriate when the error distribution is expected to be Gaussian (Chai and Draxler 2014), such as may be expected from time series data averaged over multiple model analysis base times or forecast horizons, and draws attention to model grid

cells at time steps containing larger errors. However, given that RMSE is constructed in multiple steps (sum of squared error, mean of the sum, and the square root of the sum) (Willmott and Matsuura 2005) the interpretation of the metric can oftentimes prove challenging.

All metrics were calculated on postprocessed model analysis and forecast data using Python and the NCAR Command Language (NCL) (UCAR 2019); in particular the Iris Python Library (Met Office 2010) and custom verification code under development for the Truth Python Library (Schroeter 2018).

3. Results

Interpretation of the following results is a function of the metric evaluated and the forecast parameter of interest within the study domain (Fig. 2). No single combination of either can establish a complete picture of model performance. Hence, we assess the range of metrics and parameters to develop a broader understanding of model performance. For convenience, we adopt the language conventions *strong skill*, *weak skill*, and *under-* and *over-*forecasting to communicate results. As the S1 skill score [Eq. (1)] is of the range [0, ∞], stronger skill refers to lower values, whereas weaker skill refers to higher values. Underforecasting is defined as situations where the model forecasts a meteorological value less than that of the verifying analysis; overforecasting is defined as situations where the model forecasts a value greater than that of the verifying analysis. Model biases are communicated in terms of *positive* and *negative*, referring to situations of over and underforecasting, respectively.

a. Meridional performance

We calculated the ACCESS-G NWP 2017 annual S1 skill score as a combined average of analysis times (0000, 0600, 1200, and 1800 UTC) at each forecast horizon over each latitude band of the entire global domain for both MSLP and 500-hPa height (Fig. 3). The skill profile of MSLP shows that the model is weakest at the equator and toward the poles, with the strongest skill around 50°S at all forecast horizons. The profile also appears to rapidly improve between 80° and 90°S. This increase in skill is likely due to a combination of factors, such as fewer available observations (and as a consequence increased influence from the initial model forecast background), the reduction from surface pressure to sea level for MSLP in the model, and meridional convergence. This suggests a peculiarity of the metrics rather than a rapid performance increase. Equatorial performance remains stable, albeit poor, irrespective of forecast horizon (Figs. 3e–g). Figures 3e–g show a weakening of skill toward toward 70°S, after which skill begins to
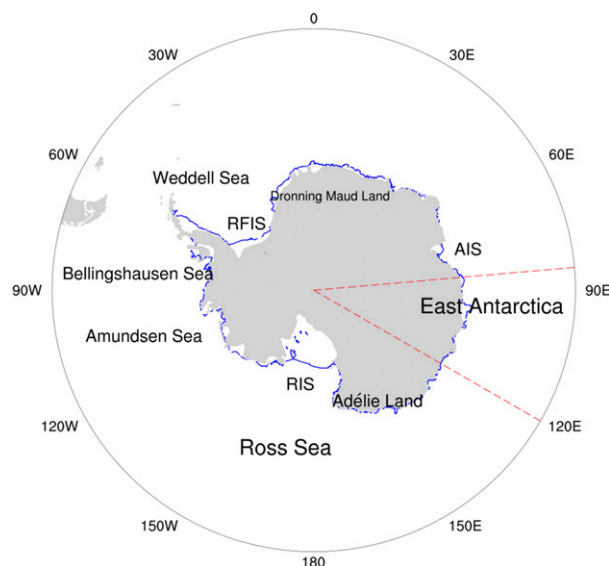
FIG. 2. Map of the study domain annotated with the regions discussed; including the Ronne Filchner Ice Shelf (RFIS), Amery Ice Shelf (AIS), and the Ross Ice Shelf (RIS). The dashed red lines indicate the location of atmospheric transects through 85° and 120°E. The blue line indicates the approximate ice-shelf edge.

improve toward the poles. Arguably, the increasing proportion of land to ocean and the derivation of MSLP over land in the model may account for some of this increase in skill. The slope either side of this inflexion point steepens with longer to forecast horizon, suggesting that skill performance becomes less stable at longer forecast horizons.

The S1 skill score of the 500-hPa geopotential height field is also weakest at the equator and toward the poles, with the strongest skill also around 50°S (Fig. 3a). Again, while equatorial performance is poor, it is consistently so at longer forecast horizons. We note that the skill profile of 500-hPa geopotential height is considerably smoother than that of MSLP, with the 500-hPa skill outperforming the MSLP skill through from the midlatitudes to approximately 70°S. This suggests potential topographic influences, as 500 hPa is positioned in the free atmosphere above the planetary boundary layer where topography is less influential. Despite topographical problems, when plotted as a function of the 12-h forecast (Figs. 3e–g) 500-hPa geopotential height degrades faster than MSLP at longer forecast horizons.

RMSE and MAE of MSLP (Figs. 4a–d) indicate decreasing performance toward the poles. Performance weakens at longer forecast horizons and is worst at around 60°S for both metrics. Using this latitude for reference, RMSE performance is approximately three times weaker at 48 h as it is at 12 h. Similarly, ME and MAE are about 2 times weaker over the same time

period. The meridional profiles of these metrics show a performance degradation toward the high latitudes, particularly to the south. The slope of this degradation steepens at longer to forecast horizon. The meridional profile of ME for MSLP does not exhibit this behavior, rather the metric tends toward zero with slight positive biases at about 30°S and south of 75°S.

The meridional profiles of surface pressure (Figs. 4e–h) follow a similar pattern to MSLP, albeit with a steeper slope to the RMSE/MAE maximum at 60°S. If we use this latitude for reference, the 48-h RMSE of surface pressure is approximately 2.5 times that of the 12-h forecast.

ME of screen temperature shows a negative model bias for much of the midlatitudes and at 80°S where the parameter is underforecast (Figs. 4i–l). This bias is reflected in both MAE and RMSE, albeit to a lesser extent due to the averaging used to produce the profile. There are RMSE and MAE maxima at 80°S, which given the coincidence with strong negative model bias (shown in the ME profile) suggests that these errors consist of greater instances of underforecasting of the parameter. Specifically, the model is forecasting temperatures that are too cold in comparison to the reference analysis with a number of outliers that contribute to an elevated RMSE, increasing error variance. As with MSLP and surface pressure, the minima and maxima of the meridional error profile of screen temperature become exaggerated at longer forecast horizons, but still exhibit similar behavior. Taking the reference latitude of 80°S, RMSE and MAE are approximately 3 and 4 times worse at 48 h than at 12 h, respectively.

The meridional error profiles of the zonal u wind component of the model show a predominantly positive (overforecast) model bias in ME over the Antarctic domain, with exceptions at 70°S and south of 80°S where the bias becomes negative (Figs. 4m-p). RMSE and MAE show a maximum about 65°S, suggesting that while ME may be tending toward negative at this latitude, that there are strong outliers that contribute to higher RMSE scores. There are potential diurnal signatures in the ME profile, with the 24- and 48-h forecasts (Figs. 4n and 4p) showing weaker negative biases than at 36 h toward the pole. Interestingly, shorter forecast horizons (Figs. 4n-o) show positive model ME biases around 50°S with slightly elevated MAE scores at this latitude. Taking a reference latitude of 65°S, the 48-h forecast shows errors three times that of the 12-h forecast. Again, the slope of weakening RMSE/MAE toward the high southern latitudes steepens with forecast horizon and the ME profile becomes more exaggerated. The error profiles of the meridional v wind component of the model (Figs. 4q-t) follow a similar pattern to the zonal wind component, with the notable
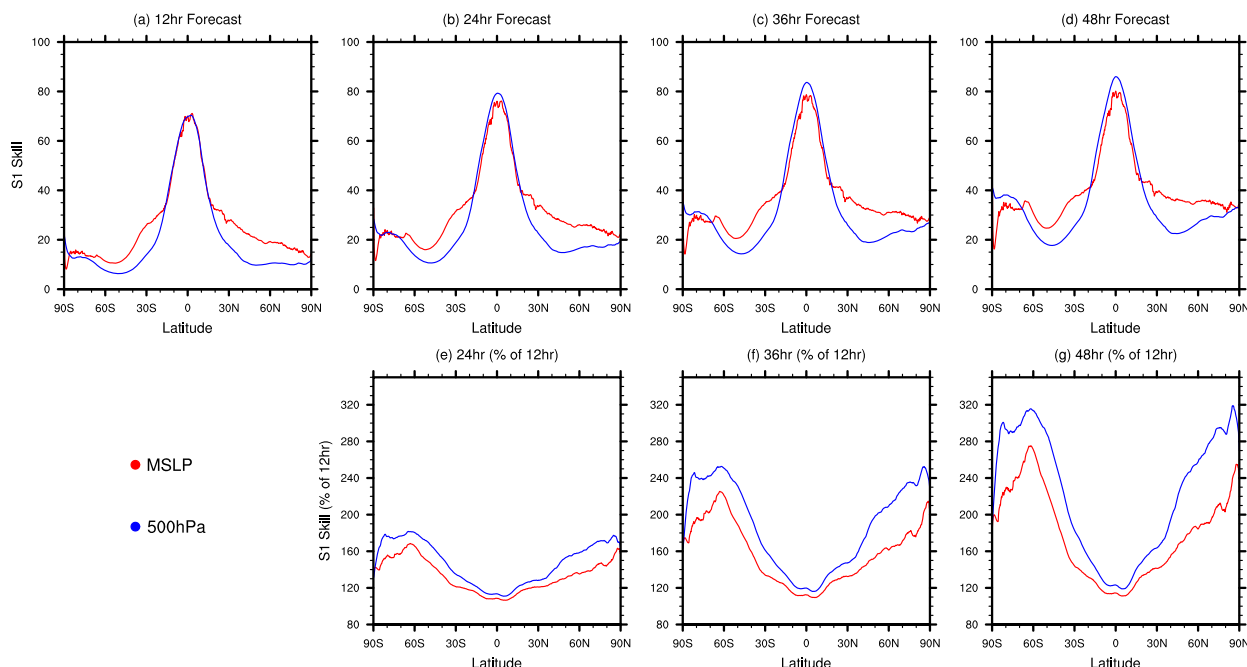
FIG. 3. ACCESS-G 2017 annual average S1 skill score as a combined average of analysis times (0000, 0600, 1200, 1800 UTC) for 12–48-h forecast horizon calculated over each latitude band (a)–(d) for MSLP and 500-hPa geopotential height and (e)–(g) depicted as a percentage of the 12-h forecast.

exception of RMSE/MAE maxima occurring at slightly lower latitudes (60°S).

The meridional error profiles of geopotential height at 500 hPa (Figs. 4u–x) show a smoother profile in RMSE and MAE with maxima around 70°S; this profile follows the steepening slope toward the poles depicted by other parameters. Oddly, the ME profile is considerably different to other parameters and metrics, with a defined negative bias across the majority of the global latitude range (60°N–60°S). A positive model ME bias occurs at 70°S at all horizons, becoming greater at longer forecast leads. The coincidence of these positive ME biases with RMSE and MAE maxima suggest that errors at these latitudes include those that are mostly positive and with strong outliers. There is inconsistent model behavior at 80°S, which remains around zero at 12, 36, and 48 h, but not at 24 h. This suggest some temporal influence (such as diurnal processes or assimilation of observations) adversely affecting the forecast at 24 h. Furthermore, this zero point in the ME profile moves farther south at longer forecast horizons before becoming positively biased toward the pole.

b. Spatial distribution of model performance

The ME performance distributions of MSLP and surface pressure (Figs. 5a–h) show positive model biases near the coast over the Ross Sea. These biases intensify at longer forecast horizons and reach magnitudes of

2 hPa or greater, particularly to the east of Adelie Land. To the west, a strong discontinuity leading to strong negative biases (also intensifying at longer forecast horizons) covers large portions of East Antarctica approaching 2-hPa divergence from the reference analyses. The Ronne Filchner Ice Shelf (RFIS) is also another site of intensifying negative biases. There are positive screen temperature ME biases approaching 1.5–2.0 K between 100° and 150°E extending to approximately 60°S (Figs. 5i–l). Negative biases of comparable magnitude (but not distribution) occur around 75°S, 80°E.

The zonal wind u component of the model yields positive biases approaching $1 \, \mathrm{m \, s^{-1}}$ over parts of Dronning Maud Land, along the coast east of the Amery Ice Shelf (AIS) and in the lee of Adelie Land, where errors increase toward $2 \, \mathrm{m \, s^{-1}}$ (Figs. 5m–p). The meridional wind v component of the model is substantially underforecast between 90° and 150°E, by more than $2 \, \mathrm{m \, s^{-1}}$ at longer forecast horizons (Figs. 5q–t).

ME of geopotential height at 500 hPa shows a strong negative biases between 0° and 150°E, where the model underforecasts the geopotential height by up to 10 m at longer forecast horizons (Figs. 5u–x). Conversely, the model overforecasts the height field over the Ross, Amundsen, Bellingshausen, and Weddell Seas; with errors penetrating inland from 120°E to 90°W. These positive biases also increase at longer forecasts.
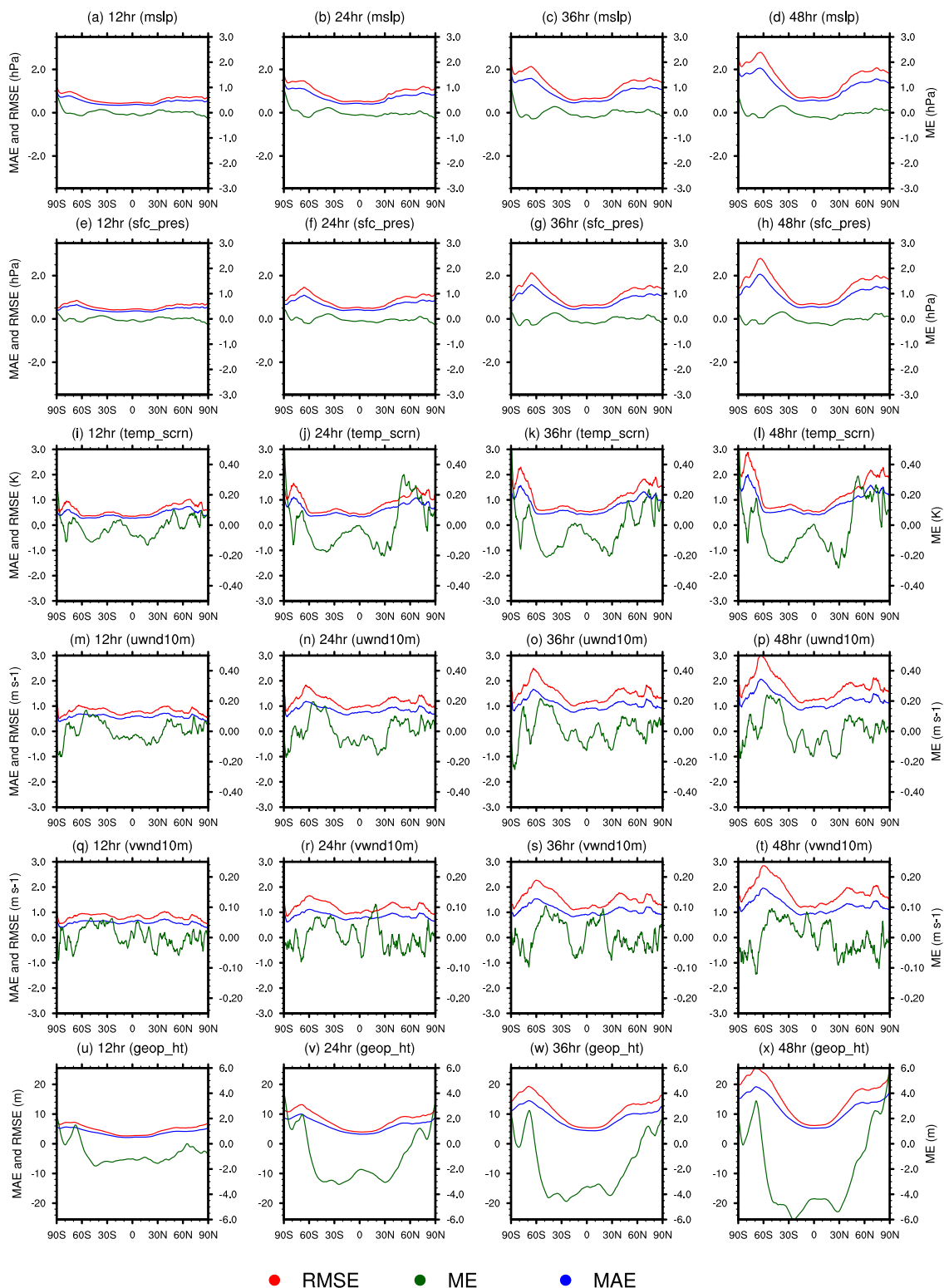
FIG. 4. Meridional error profiles for (a)–(d) MSLP, (e)–(h) surface pressure, (i)–(l) screen temperature, (m)–(p) zonal wind, (q)–(t) meridional wind, and (u)–(x) 500-hPa geopotential height for the combined average analysis base times (0000, 0600, 1200, 1800 UTC) at each forecast horizon (12, 24, 36, 48 h).
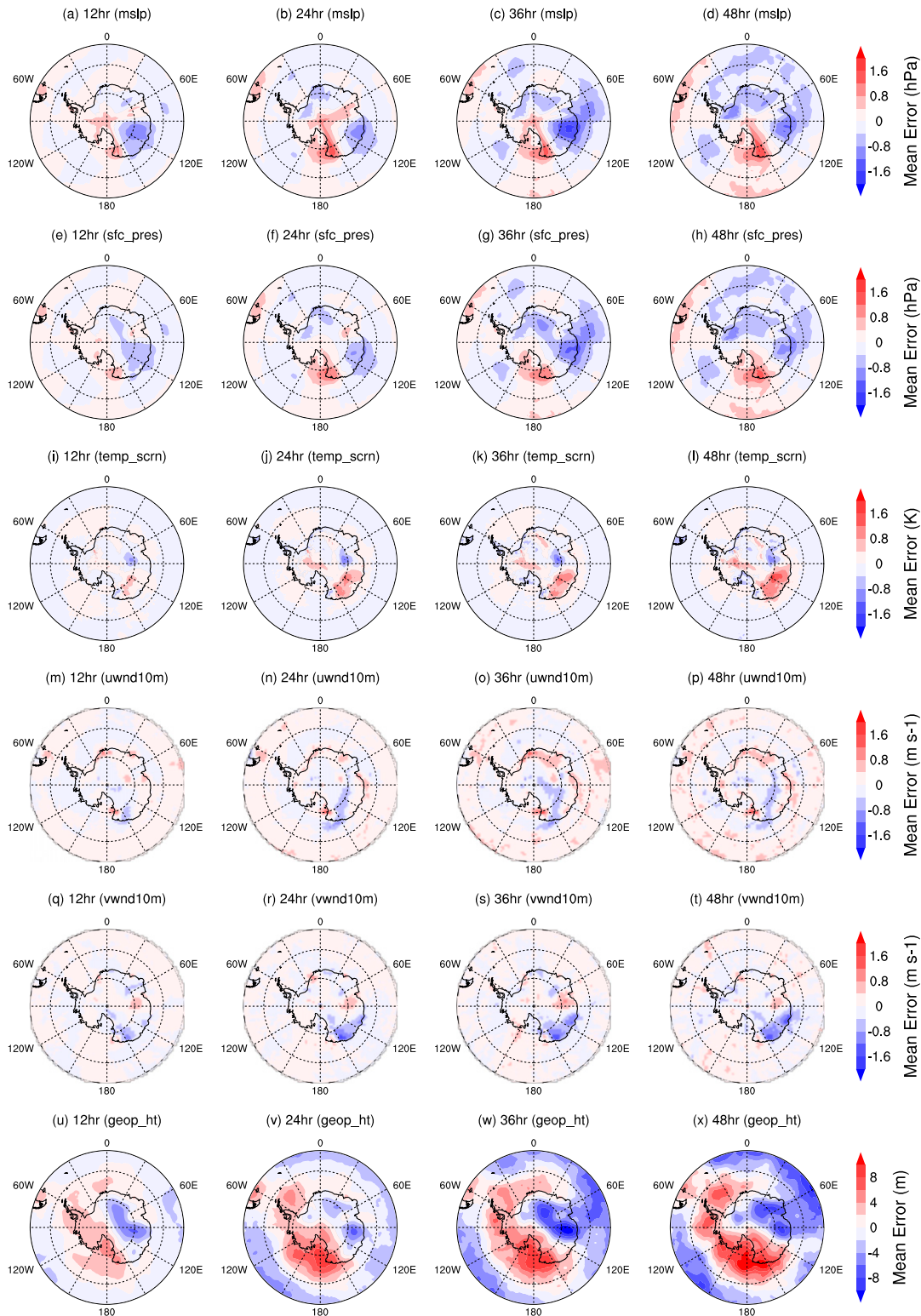
FIG. 5. ACCESS-G 2017 annual average mean error (through time) as a combined average of analysis times (0000, 0600, 1200, 1800 UTC) for 12–48-h forecast horizon for (a)–(d) MSLP, (e)–(h) surface pressure, (i)–(l) screen temperature, (m)–(p) zonal wind, (q)–(t) meridional wind, and (u)–(x) 500-hPa geopotential height. The inner latitude reference circle is 60°S.
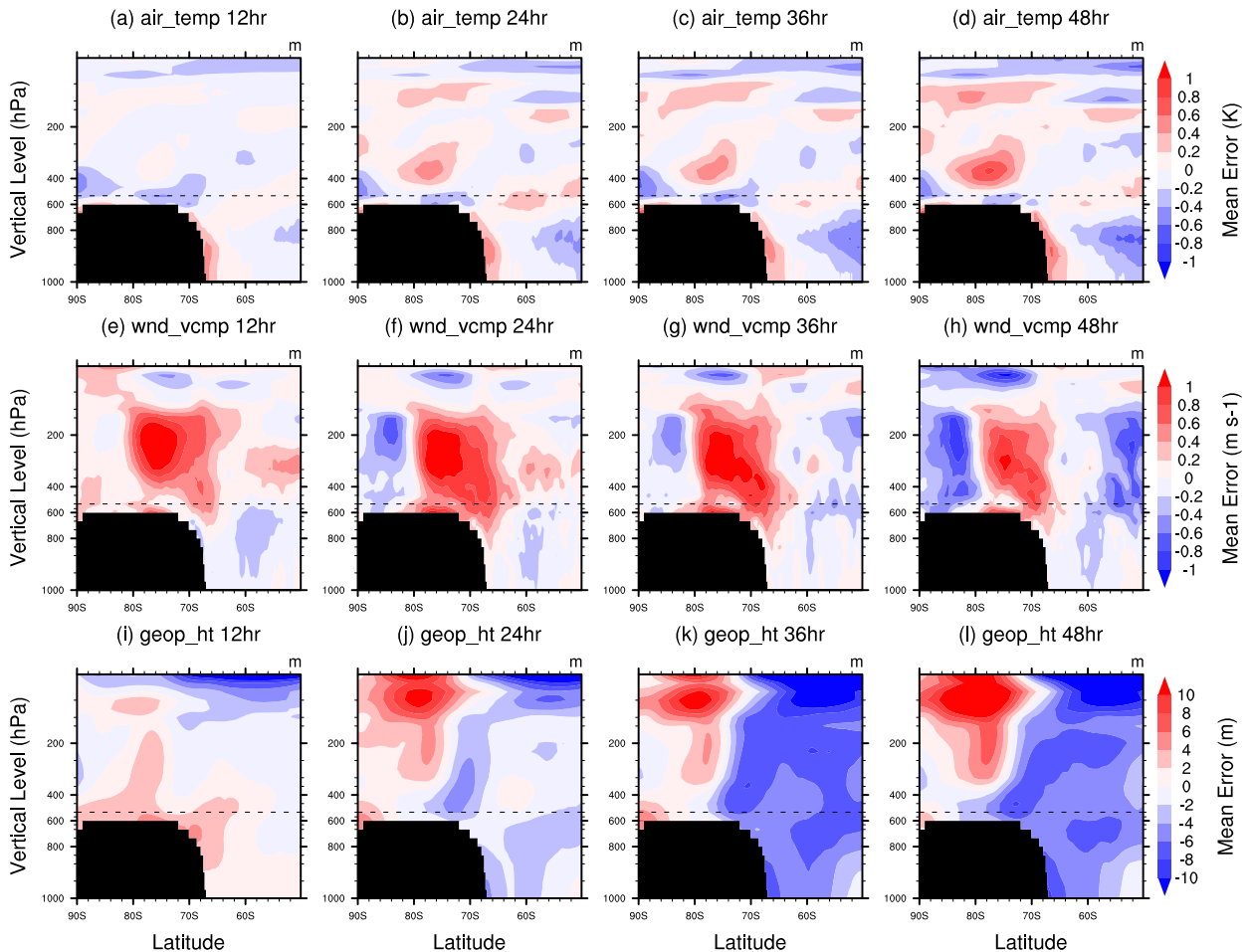
FIG. 6. ACCESS-G 2017 atmospheric transects through 85°E for the 0000 UTC analysis at each forecast horizon for mean error in (a)–(d) air temperature, (e)–(h) meridional wind, and (i)–(l) geopotential height. The dashed line is the theoretical 500-hPa isobaric surface.

c. Atmospheric transects

To better understand the nature of model biases throughout the atmospheric column, we calculated vertical atmospheric transects through the longitudes of regions presenting strong positive and negative error behavior. We have plotted transects of theses error profiles through 85° and 120°E, respectively, to assess regions of larger positive and negative model biases.

The model exhibits a positive (warm) surface temperature bias through 85°E over sloping topography at all forecast horizons, becoming stronger as forecast length increases (Figs. 6a–d). Similarly, there is also a warm bias in the mid- and upper atmosphere over land that also intensifies at longer forecasts. In contrast, there is a negative (cold) model temperature bias over the ocean and close to the surface over smooth topography, as well as a poleward cool bias that contracts poleward at longer forecasts and is delineated by the theoretical 500-hPa

surface above the planetary boundary layer (PBL) (Figs. 6a–d). A positive (fast) meridional wind bias is also shown between 65° and 80°S, which is delineated by a negative (slow) bias to the north and south (Figs. 6e–h). These biases extend through almost the full atmospheric column and overforecast winds are likely driven by the high to low pressure gradient illustrated in Figs. 6i–l.

A positive (elevated) bias exists in geopotential height at the surface and in the midatmosphere through 85°E at 12 h (Fig. 6i). However, this positive bias detaches from the surface at longer forecasts, when negative (depressed) biases over the ocean begin to dominate at the surface and throughout the atmospheric column over the ocean (Figs. 6j–l).

Figure 7 illustrates the vertical error profile of an atmospheric transect through 120°E, where error extremes have been noted previously. A positive (warm) surface bias is present in the model, particularly over steep topography and toward the coast through 120°E (Figs. 7a–d). This is
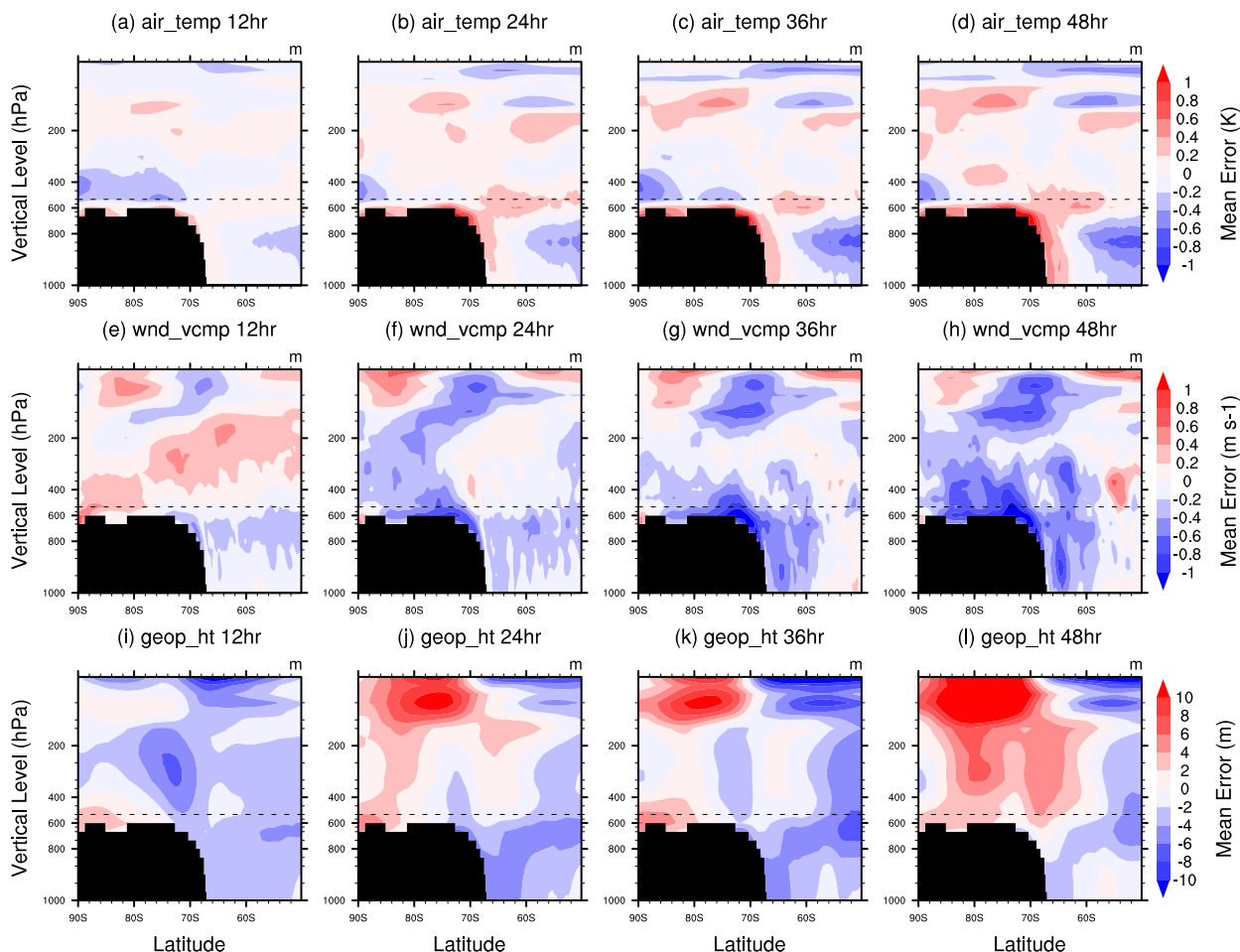
FIG. 7. ACCESS-G 2017 atmospheric transects through 120°E for the 0000 UTC analysis at each forecast horizon for mean error in (a)–(d) air temperature, (e)–(h) meridional wind, and (i)–(l) geopotential height. The dashed line is the theoretical 500-hPa isobaric surface.

coincident with a negative (slow) meridional wind bias in the same region, which is consistent out to longer forecast lengths (Figs. 7e–h). These warm biases appear to advect farther to the north aloft of cold ocean biases as the forecast horizon increases, remaining attached to the land at 24 and 48 h (Figs. 7b,d). As with the transects through 85°E (Fig. 6), negative temperature biases are delineated by the theoretical PBL at 500 hPa, contracting poleward and with a subtle diurnality covering a greater meridional range of the continent at 12 and 36 h (Figs. 7a–d).

The model overforecasts meridional winds over land and aloft of negative ocean wind biases at 12 h; however, these biases change sign at longer forecast horizons (Figs. 7e–h). Ocean biases remain negative and increase in intensity from 12 to 48 h. The 500-hPa geopotential surface is under forecast (too low) in the model over the ocean through 120°E at all forecast horizons (Figs. 7i–l). There are potential diurnal influences across the forecast horizons examined, with 24- and 48-h forecast connecting

surface and upper atmosphere positive biases throughout the vertical column more substantially than at 12 and 36 h. Given that the time series was averaged over an entire year, seasonal analyses not covered in this study would yield greater insight into diurnal influences.

4. Discussion

The performance of ACCESS-G NWP weakens toward the high southern latitudes, most notably toward the Antarctic continent. This behavior is consistent under a range of different performance metrics. The S1 skill profile of the model at each latitude shows not only that model performance is reduced toward the pole, but it reduces at a greater rate as the forecast horizon increases (Fig. 3). While there may be peculiarities with the S1 metric (such as sensitivities to grid structure and resolution), this behavior is also present to varying degrees in the meridional profiles of RMSE, MAE and ME

for the additional meteorological parameters examined in this study (Fig. 4).

It should be generally noted that the utility of the S1 skill score is limited in this study, as meridional convergence and the lack of observations available for data assimilation (see Puri et al. 2013; Australian Bureau of Meteorology 2016) yield an unrepresentative measure of model performance; this is also the case with the other metrics examined in this study. Arguably, the lack of observations has substantial influence in the high southern latitudes with a dearth of surface observations and satellite measurements rejected or simply unavailable for data assimilation, particularly during winter. Conversely, the high NH latitudes are comparatively better sampled than the SH. This has important implications for self-verification, whereby the model analysis will not deviate substantially from the model background (the prior forecast). Thus, in the absence of observations the model will verify against itself and appear artificially skilful.

There is an historical legacy behind the use of S1 skill and it remains a useful skill measure for SH mid- to high-latitude weather driven by horizontal gradients associated with baroclinicity. However, it must be considered in concert with other measures when evaluating model performance at the pole. Observational coverage is challenging in the Antarctic; with surface and satellite instruments that are spatially sparse, temporally intermittent (with some platforms only operating seasonally) and AWSs subject to occasional relocation as needed by base operations. Furthermore, while there is an array of AWSs distributed on and around the continent (Fig. 1) not all of the stations available are actually assimilated into ACCESS-G (see Puri et al. 2013; Australian Bureau of Meteorology 2016).

The ACCESS-G NWP model exhibits persistent negative surface pressure and MSLP biases over large parts of the continent, the strongest of which occur at the 36-h forecast over East Antarctica between 0° and 120°E. Similarly, the 500-hPa geopotential height field over the same region is systemically underforecast, with the isobaric surface approaching 10 m below the reference analysis. These parameters appear to be linked, with lower surface pressures expressed throughout the vertical column via a depression of the 500-hPa isobaric surface (Fig. 5). For context, 2017 featured positive surface pressure anomalies over the ocean in several of the regions presenting negative biases for much of the year (Clem et al. 2018). These pressure anomalies were characterized by a pronounced zonal wave-3 (ZW3) pattern that emerged in June–September and featured ridges across 50°S at 90°E, 150°E, and 30°W. The effect of these anomalies is subtle and only observable in 12–24-h MSLP and surface pressure ME fields (Figs. 5a,b), tending toward a zonal wave-1 (ZW1) pattern at longer forecast horizons with an error ridge along 150°E (Figs. 5c,d). It is possible that this strong zonal wave-3 pattern observed in 2017 impresses itself upon the errors within the model. However, it is likely that the errors associated with the Adelie Land trough and Ross Sea ridge are a linked 1-wave pattern, possibly as a consequence of the model's inability to correctly simulate atmospheric drainage over Adelie Land, with the associated errors propagating eastward as an atmospheric wave. As global NWP models are noted to exhibit sensitivities to surface and planetary boundary layer initial conditions (Powers et al. 2012), errors at the surface are arguably propagated upward. Thus, an improved representation of physical processes at the surface more suited to the unique Antarctic environment will likely yield improvements aloft.

Positive surface and MSLP biases near Adelie Land may again be associated with surface pressure anomalies observed in 2017 (Clem et al. 2018), or with cyclonic activity in the area. Furthermore, these biases may be associated with the large temperature gradient brought about by katabatic outflow from the elevated East Antarctic topography, which frequently develops into a low in the region (Chen et al. 2014; Bromwich et al. 2011). Bromwich et al. (2011) describe the processes of cyclogenesis in this region as both secondary and lee cyclogenesis whereby dissipating synoptic-scale cyclones to the west interact with the Adelie katabatic jet to spin up the secondary development of cyclones. This is expressed at the surface as surface/MSLP minima observable in Figs. 5a–h. Similarly, if the model does not capture the cyclonic activity in this area, isobaric surfaces throughout the vertical column would also be more elevated that the reference analysis, as shown in Figs. 5u–x. A contributing factor to this error behavior is the underforecasting of meridional winds over western Adelie Land (Figs. 5q–t), whereby a weakened representation of katabatic outflow would fail to reach sufficient momentum to create a closed circulation and trigger cyclogenesis. The forces driving weaker modeled winds in ACCESS-G NWP are not yet fully understood; however, Orr et al. (2014) found strong wind events in the Unified Model (which is the atmospheric core of ACCESS-G) to be sensitive to both horizontal resolution (especially at the coast) and turbulent mixing under stable conditions. As such, future studies should investigate these areas to improve model development.

Results here suggest that katabatic outflow from the Adelie Land coast is underrepresented in the ACCESS-G NWP model (Figs. 5q–t), and is likely influenced by positive temperature biases at the surface illustrated in

Fig. 7. These positive temperature biases may be associated with below-average temperatures across the continent and east (west) of midlatitude ridges (troughs) (Clem et al. 2018), poor observational sampling, or as a result of suboptimal model parameterization. However, the proximity of these temperature biases to troughs is also favorable for storm development through enhanced baroclinic instability (Chen et al. 2014), further emphasizing the importance of accurately modeled temperatures throughout the atmospheric column.

As katabatic flow is driven by both temperature and topography, a warm bias of the former leads to less resultant downslope movement of air and subsequently slower model wind speeds. Strong surface temperature inversions over ice-covered terrain, which may be much cooler than the air aloft (Hines and Bromwich 2008), could be addressed by adjusting the model's radiative and thermal properties over the Antarctic continent, such as treating upward longwave flux as a function of skin temperature (Hines and Bromwich 2008). Similarly, modifying the thermal conductivity of permanent snow/ice surfaces as a function of empirical snow density (Yen 1981) could be investigated for the Antarctic environment, as would a SH-focused snow analysis (i.e., Pullen et al. 2011).

5. Conclusions

This study has investigated the performance of ACCESS-G NWP over the high southern latitudes where the performance of the model was found to degrade toward the poles, at a rate proportional to forecast horizon. This behavior was diagnosed by several performance metrics. Evaluation of model error both spatially and vertically suggest boundary layer parameterization, initial conditions and associated physical processes may be contributing factors in the error behavior of the region, as could the anomalous surface pressure and temperature behavior observed in 2017 (Clem et al. 2018). Many of these biases are interrelated, coalescing into regional biases such as the combination of warm surface biases, weak model winds and positive surface pressure biases that inadequately represent cyclonic activity around the Adelie Land coast.

The biases examined in this paper could be addressed through an improved representation of the physical processes governing model initialization and boundary layer parameterization over the unique Antarctic region (see Tastula and Vihma 2011; Powers et al. 2003), which have been shown to be sensitive to initial conditions over frozen surfaces (Hines et al. 2011). Improving model performance in the region would likely yield improved model forecast guidance to those operating in the region. However, this is largely speculative and further model experimentation is required. As such, future ACCESS-G development should focus on better representation of Antarctic processes to improve overall model performance.

Additional observations made available for data assimilation would also likely yield improvements to the model initial conditions, as may increased model resolution. Given the logistical and financial challenges of installing and maintaining in situ observing systems, this requires modelers to make greater use of remotely sensed and satellite observations for data assimilation and verification purposes (Casati et al. 2017).

We acknowledge the limitations of this study, specifically the use of model analysis as a reference dataset for verification and the use of a single year of data. Given the development schedules of the ACCESS family of models, data from 2017 were the most consistent and complete, across a full calendar year. Ideally, a longer time series and additional observational data would provide additional context around model performance, as would a seasonally focused study.

REFERENCES

Abel, S. J., and B. J. Shipway, 2007: A comparison of cloud-resolving model simulations of trade wind cumulus with aircraft observations taken during RICO. *Quart. J. Roy. Meteor. Soc.*, **133**, 781–794, https://doi.org/10.1002/qj.55.

Antarctic Meteorological Research Center, 2018: Automatic Weather Stations— Antarctica 2018. Antarctic Meteorological Research Center, accessed 24 August 2018, https://amrc.ssec.wisc.edu/aws/documents/2018_AWS_Sites_ALL_03_29_2018.pdf.

Australian Bureau of Meteorology, 2016: BNOC Operations Bulletin Number 105: APS2 upgrade to the ACCESS-G numerical weather prediction system. Tech. Rep. 105, Bureau National Operations Centre, 32 pp., http://www.bom.gov.au/australia/charts/bulletins/APOB105.pdf.

Bauer, P., A. Thorpe, and G. Brunet, 2015: The quiet revolution of numerical weather prediction. *Nature*, **525**, 47–55, https://doi.org/10.1038/nature14956.

Beggs, H., 2008: GAMSSA–A new Global Australian Multi-Sensor SST Analysis. *Proc. Ninth GHRSSTûPP Science Team Meeting*, Perros-Guirec, France, GHRSST, 9–13, https://www.ghrsst.org/meetings/9th-international-ghrsst-science-team-meeting-ghrsst-ix/.

Bengtsson, L., 1991: Advances and prospects in numerical weather prediction. *Quart. J. Roy. Meteor. Soc.*, **117**, 855–902, https://doi.org/10.1002/qj.49711750102.

Best, M. J., and Coauthors, 2011: The Joint UK Land Environment Simulator (JULES), model description—Part 1: Energy and water fluxes. *Geosci. Model Dev.*, **4**, 677–699, https://doi.org/10.5194/gmd-4-677-2011.

Boutle, I. A., and C. J. Morcrette, 2010: Parametrization of area cloud fraction. *Atmos. Sci. Lett.*, **11**, 283–289, https://doi.org/10.1002/asl.293.

Bracegirdle, T. J., and G. J. Marshall, 2012: The reliability of Antarctic tropospheric pressure and temperature in the latest global reanalyses. *J. Climate*, **25**, 7138–7146, https://doi.org/10.1175/JCLI-D-11-00685.1.

Bromwich, D. H., A. J. Monaghan, K. W. Manning, and J. G. Powers, 2005: Real-time forecasting for the Antarctic: An evaluation of the Antarctic Mesoscale Prediction System (AMPS). *Mon. Wea. Rev.*, **133**, 579–603, https://doi.org/10.1175/MWR-2881.1.

——, D. F. Steinhoff, I. Simmonds, K. Keay, and R. L. Fogt, 2011: Climatological aspects of cyclogenesis near Adélie Land Antarctica. *Tellus*, **63A**, 921–938, https://doi.org/10.1111/j.1600-0870.2011.00537.x.

Casati, B., T. Haiden, B. Brown, P. Nurmi, and J.-F. Lemieux, 2017: Verification of environmental prediction in polar regions: Recommendations for the Year of Polar Prediction. Tech. Rep. WWRP 2017-1, 44 pp., https://www.polarprediction.net/fileadmin/user_upload/www.polarprediction.net/Home/Organization/Task_Teams/Verification/Casati.YOPPverif.final2017.pdf.

Chai, T., and R. R. Draxler, 2014: Root mean square error (RMSE) or mean absolute error (MAE)?—Arguments against avoiding RMSE in the literature. *Geosci. Model Dev.*, **7**, 1247–1250, https://doi.org/10.5194/gmd-7-1247-2014.

Chen, S.-Y., T.-K. Wee, Y.-H. Kuo, and D. H. Bromwich, 2014: An impact assessment of GPS radio occultation data on prediction of a rapidly developing cyclone over the Southern Ocean. *Mon. Wea. Rev.*, **142**, 4187–4206, https://doi.org/10.1175/MWR-D-14-00024.1.

Clem, K. R., S. Barreira, R. L. Fogt, S. Colwell, C. Costanza, L. M. Keller, and M. A. Lazzara, 2018: Atmospheric circulation and surface observations [in "State of the Climate in 2017"]. *Bull. Amer. Meteor. Soc.*, **99** (8), S176–S179.

Comiso, J. C., 2000: Variability and trends in Antarctic surface temperatures from in situ and satellite infrared measurements. *J. Climate*, **13**, 1674–1696, https://doi.org/10.1175/1520-0442(2000)013<1674:VATIAS>2.0.CO;2.

Connolley, W. M., and S. A. Harangozo, 2001: A comparison of five numerical weather prediction analysis climatologies in southern high latitudes. *J. Climate*, **14**, 30–44, https://doi.org/10.1175/1520-0442(2001)014<0030:ACOFNW>2.0.CO;2.

Crocker, R., and M. Mittermaier, 2013: Exploratory use of a satellite cloud mask to verify NWP models. *Meteor. Appl.*, **20**, 197–205, https://doi.org/10.1002/met.1384.

Cullen, M., 1993: The unified forecast/climate model. *Meteor. Mag.*, **122** (1449), 81–94.

Davies, T., M. J. Cullen, A. J. Malcolm, M. H. Mawson, A. Staniforth, A. A. White, and N. Wood, 2005: A new dynamical core for the Met Office's global and regional modelling of the atmosphere. *Quart. J. Roy. Meteor. Soc.*, **131**, 1759–1782, https://doi.org/10.1256/qj.04.101.

Ebert, E., and Coauthors, 2013: Progress and challenges in forecast verification. *Meteor. Appl.*, **20**, 130–139, https://doi.org/10.1002/met.1392.

Edwards, J. M., and A. Slingo, 1996: Studies with a flexible new radiation code. I: Choosing a configuration for a large-scale model. *Quart. J. Roy. Meteor. Soc.*, **122**, 689–719, https://doi.org/10.1002/qj.49712253107.

Eerola, K., 2013: Twenty-one years of verification from the HIRLAM NWP system. *Wea. Forecasting*, **28**, 270–285, https://doi.org/10.1175/WAF-D-12-00068.1.

Goessling, H. F., and Coauthors, 2016: Paving the way for the Year of Polar Prediction. *Bull. Amer. Meteor. Soc.*, **97**, ES85–ES88, https://doi.org/10.1175/BAMS-D-15-00270.1.

Gregory, D., and P. R. Rowntree, 1990: A mass flux convection scheme with representation of cloud ensemble characteristics and stability-dependent closure. *Mon. Wea. Rev.*, **118**, 1483–1506, https://doi.org/10.1175/1520-0493(1990)118<1483:AMFCSW>2.0.CO;2.

Hines, K. M., and D. H. Bromwich, 2008: Development and testing of Polar Weather Research and Forecasting (WRF) Model. Part I: Greenland Ice Sheet meteorology. *Mon. Wea. Rev.*, **136**, 1971–1989, https://doi.org/10.1175/2007MWR2112.1.

——, ——, L. S. Bai, M. Barlage, and A. G. Slater, 2011: Development and testing of Polar WRF. Part III: Arctic land. *J. Climate*, **24**, 26–48, https://doi.org/10.1175/2010JCLI3460.1.

Jolliffe, I. T., and D. B. Stephenson, Eds., 2012: *Forecast Verification: A Practitioner's Guide in Atmospheric Science.* 2nd ed. Wiley, 292 pp.

Jung, T., and M. Matsueda, 2016: Verification of global numerical weather forecasting systems in polar regions using TIGGE data. *Quart. J. Roy. Meteor. Soc.*, **142**, 574–582, https://doi.org/10.1002/qj.2437.

Kaimal, J. C., and J. J. Finnigan, 1994: *Atmospheric Boundary Layer Flows: Their Structure and Measurement*. Oxford University Press, 304 pp.

Lazzara, M. A., G. A. Weidner, L. M. Keller, J. E. Thom, and J. J. Cassano, 2012: Antarctic Automatic Weather Station Program: 30 Years of polar observation. *Bull. Amer. Meteor. Soc.*, **93**, 1519–1537, https://doi.org/10.1175/BAMS-D-11-00015.1.

Met Office, 2010: Iris: A Python library for analysing and visualising meteorological and oceanographic data sets. Met Office, accessed 3 April 2017, https://scitools.org.uk/.

NOAA, 2018: Pressure altitude. NOAA, accessed 10 June 2017, https://www.weather.gov/media/epz/wxcalc/pressureAltitude.pdf.

Orr, A., T. Phillips, S. Webster, A. Elvidge, M. Weeks, S. Hosking, and J. Turner, 2014: Met Office Unified Model high-resolution simulations of a strong wind event in Antarctica. *Quart. J. Roy. Meteor. Soc.*, **140**, 2287–2297, https://doi.org/10.1002/qj.2296.

Pendlebury, S. F., N. D. Adams, T. L. Hart, and J. Turner, 2003: Numerical weather prediction model performance over high southern latitudes. *Mon. Wea. Rev.*, **131**, 335–353, https://doi.org/10.1175/1520-0493(2003)131<0335:NWPMPO>2.0.CO;2.

Powers, J. G., A. J. Monaghan, A. M. Cayette, D. H. Bromwich, Y.-H. Kuo, and K. W. Manning, 2003: Real-time mesoscale modeling over Antarctica: The Antarctic Mesoscale Prediction System. *Bull. Amer. Meteor. Soc.*, **84**, 1533–1546, https://doi.org/10.1175/BAMS-84-11-1533.

——, K. W. Manning, D. H. Bromwich, J. J. Cassano, and A. M. Cayette, 2012: A decade of Antarctic science support through AMPS. *Bull. Amer. Meteor. Soc.*, **93**, 1699–1712, https://doi.org/10.1175/BAMS-D-11-00186.1.

Pullen, S., C. Jones, and G. Rooney, 2011: Using satellite-derived snow cover data to implement a snow analysis in the Met Office global NWP model. *J. Appl. Meteor. Climatol.*, **50**, 958–973, https://doi.org/10.1175/2010JAMC2527.1.

Puri, K., and Coauthors, 2013: Implementation of the initial ACCESS numerical weather prediction system. *Aust. Meteor. Oceanogr. J.*, **63** (2), 265–284.

Rawlins, F., S. P. Ballard, K. J. Bovis, A. M. Clayton, D. Li, G. W. Inverarity, A. C. Lorenc, and T. J. Payne, 2007: The Met Office global four-dimensional variational data assimilation scheme. *Quart. J. Roy. Meteor. Soc.*, **133**, 347–362, https://doi.org/10.1002/qj.32.

Schroeter, B. J. E., 2018: Truth: A Python library for the verification of Earth System Data. Github, accessed 4 August 2017, https://github.com/bschroeter/truth.

Skamarock, W. C., and Coauthors, 2008: A description of the Advanced Research WRF version 3. NCAR Tech. Note NCAR/TN-475+STR, 113 pp., https://doi.org/10.5065/D68S4MVH.

Tastula, E.-M., and T. Vihma, 2011: WRF Model experiments on the Antarctic atmosphere in winter. *Mon. Wea. Rev.*, **139**, 1279–1291, https://doi.org/10.1175/2010MWR3478.1.

Teweles, S., and H. Wobus, 1954: Verification of prognostic charts. *Bull. Amer. Meteor. Soc.*, **35**, 455–463, https://doi.org/10.1175/1520-0477-35.10.455.

Thompson, J. C., and G. M. Carter, 1972: On some characteristics of the S1 score. *J. Appl. Meteor.*, **11**, 1384–1385, https://doi.org/10.1175/1520-0450(1972)011<1384:OSCOTS>2.0.CO;2.

UCAR, 2019: NCAR Command Language (version 6.5.0). UCAR/NCAR/CISL/TDD, accessed 26 July 2018, http://dx.doi.org/10.5065/D6WD3XH5, http://www.ncl.ucar.edu/.

Walton, D. W., Ed., 2013: *Antarctica: Global Science from a Frozen Continent.* Cambridge University Press, 342 pp.

Wilks, D. S., 2011: *Statistical Methods in the Atmospheric Sciences.* 3rd ed. International Geophysics Series, Vol. 100, Academic Press, 704 pp.

Willmott, C. J., and K. Matsuura, 2005: Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Res.*, **30**, 79–82, https://doi.org/10.3354/cr030079.

Wilson, D. R., and S. P. Ballard, 1999: A microphysically based precipitation scheme for the UK Meteorological Office Unified Model. *Quart. J. Roy. Meteor. Soc.*, **125**, 1607–1636, https://doi.org/10.1002/qj.49712555707.

World Meteorological Organization, 2015: Manual on the global data-processing and forecasting system: Volume I—Global aspects. Attachment 11.7, World Meteorological Organization, 36–40.

Wu, X., 2015: Quarterly numerical weather prediction model performance summary—July to September 2015. *J. South. Hemisphere Earth Syst. Sci.*, **65** (3–4), 434–437.

——, 2016: Quarterly numerical weather prediction model performance summary—October to December 2015. *J. South. Hemisphere Earth Syst. Sci.*, **66**, 90–93, https://doi.org/10.22499/3.6601.008.

Yen, Y.-C., 1981: Review of thermal properties of snow, ice and sea ice. Tech. Rep. CRREL Rep. 81–10, U.S. Army Corps. of Engineers, 37 pp., https://apps.dtic.mil/dtic/tr/fulltext/u2/a103734.pdf.