Observed and Modeled Mountain Waves from the Surface to the Mesosphere Near the Drake Passage

Christopher G. Kruse,[a] M. Joan Alexander,[a] Lars Hoffmann,[b] Annelize van Niekerk,[c] Inna Polichtchouk,[d] Julio T. Bacmeister,[e] Laura Holt,[a] Riwal Plougonven,[f] Petr Šácha,[g,h] Corwin Wright,[i] Kaoru Sato,[j] Ryosuke Shibuya,[k] Sonja Gisinger,[l] Manfred Ern,[m] Catrin I. Meyer,[b] and Olaf Stein[b]

[a] *NorthWest Research Associates, Boulder, Colorado, USA*

[b] *Jülich Supercomputing Centre, Forschungszentrum Jülich, Jülich, Germany*

[c] *Met Office, Exeter, UK*

[d] *ECMWF, Reading, UK*

[e] *Climate and Global Dynamics Laboratory, NCAR, Boulder, Colorado, USA*

[f] *Laboratoire de Météorologie Dynamique, Ecole Polytechnique, Palaiseau, France*

[g] *Department of Atmospheric Physics, Faculty of Mathematics and Physics, Charles University, Prague, Czech Republic*

[h] *Institute of Meteorology and Climatology, University of Natural Resources and Life Sciences, Vienna (BOKU)*

[i] *Centre for Space, Atmospheric and Oceanic Science, University of Bath, Bath, UK*

[j] *Department of Earth and Planetary Science, University of Tokyo, Tokyo, Japan*

[k] *Atmosphere and Ocean Research Institute, The University of Tokyo, Kashiwa, Japan*

[l] *Institute of Atmospheric Physics, Deutsches Zentrum für Luft- und Raumfahrt, Oberpfaffenhofen, Germany*

[m] *Institut für Energie- und Klimaforschung - Stratosphäre (IEK-7), Forschungszentrum Jülich, Jülich, Germany*

Corresponding author: Christopher G. Kruse, ckruse@nwra.com

ABSTRACT: Four state-of-the-science numerical weather prediction (NWP) models were used to perform mountain wave- (MW) resolving hind-casts over the Drake Passage of a 10-day period in 2010 with numerous observed MW cases. The Integrated Forecast System (IFS) and the Icosahedral Nonhydrostatic (ICON) model were run at $\Delta x \approx 9$ and 13 km globally. The Weather Research and Forecasting (WRF) model and the Met Office Unified Model (UM) were both configured with a $\Delta x = 3$ km regional domain. All domains had tops near 1 Pa ($z \approx 80$ km). These deep domains allowed *quantitative* validation against Atmospheric InfraRed Sounder (AIRS) observations, accounting for observation time, viewing geometry, and radiative transfer.

All models reproduced observed middle-atmosphere MWs with remarkable skill. Increased horizontal resolution improved validations. Still, all models underrepresented observed MW amplitudes, even after accounting for model effective resolution and instrument noise, suggesting even at $\Delta x \approx 3$ km resolution, small-scale MWs are under-resolved and/or over-diffused. MW drag parameterizations are still necessary in NWP models at current operational resolutions of $\Delta x \approx 10$ km. Upper GW sponge layers in the operationally configured models significantly, artificially reduced MW amplitudes in the upper stratosphere and mesosphere. In the IFS, parameterized GW drags partly compensated this deficiency, but still, total drags were ≈ 6 time smaller than that resolved at $\Delta x \approx 3$ km. Meridionally propagating MWs significantly enhance zonal drag over the Drake Passage. Interestingly, drag associated with meridional fluxes of zonal momentum (i.e. $\overline{u'v'}$) were important; not accounting for these terms results in a drag in the wrong direction at and below the polar night jet.

SIGNIFICANCE STATEMENT: This study had three purposes: to quantitatively evaluate how well four state-of-the-science weather models could reproduce observed mountain waves (MWs) in the middle atmosphere, to compare the simulated MWs within the models, and to quantitatively evaluate two MW parameterizations in a widely used climate model. These models reproduced observed MWs with remarkable skill. Still, MW parameterizations are necessary in current $\Delta x \approx 10$-km resolution global weather models. Even $\Delta x \approx 3$-km resolution does not appear to be high enough to represent all momentum-fluxing MW scales. Meridionally-propagating MWs can significantly influence zonal winds over the Drake Passage. Parameterizations that handle horizontal propagation may need to consider horizontal fluxes of horizontal momentum in order to get the direction of their forcing correct.

1. Introduction

Orographic gravity waves (GWs), or mountain waves (MWs), are internal GWs within the atmosphere forced by stratified flow over topography. MW generation is one way the atmosphere exchanges momentum with the Earth. Positive and negative pressure perturbations upstream and downstream of a mountain, respectively, exert a net force by the atmosphere on the mountain. An equal and opposite force is exerted by the mountain on the atmosphere. Often, a MW is generated, which propagates the force by the mountain on the atmosphere upward. This force is ultimately exerted wherever the MW breaks, which can occur from the troposphere to the lower portion of the thermosphere (e.g. Fritts and Alexander 2003; Fritts et al. 2016).

MWs have a variety of horizontal scales, ranging from $\lambda_h \approx 10$ km to hundreds of kilometers, depending on the terrain scales that force them. In the DEEPWAVE field campaign (Fritts et al. 2016), aircraft-observed MWs with horizontal scales larger and smaller than 50 km were found to be equally important, fluxing similar amounts of momentum (Smith and Kruse 2017). Given the effective resolutions of current numerical weather prediction ($6\Delta x \approx 60$ km) and climate ($6\Delta x \approx 600$ km) models, an important portion of the MW spectrum is under- or unresolved and therefore needs to be parameterized. GW drag parameterizations were first formulated in the early 1980s (e.g. Lindzen 1981; Holton 1983), and the addition of MW drag (MWD) parameterizations did improve weather and climate models (e.g. Palmer et al. 1986; McFarlane 1987; Miller et al. 1989) and were widely adopted thereafter. The momentum deposited by MW breaking is important at nearly all

levels of the atmosphere, though, its importance with respect to the atmosphere's zonally-averaged momentum budget generally maximizes in the upper stratosphere and lower mesosphere, at least according to MWD parameterizations (e.g. Kruse 2018, 2020). For an overview of the ≈100-year history of MW research, see Smith (2019).

Despite the rich literature and knowledge gained over the last century, current MWD parameterizations, and GW drag parameterizations in general, are still highly simplified and not well-constrained. Most current MWD parameterizations estimate one or two 2-D source MWs based on sub-grid-scale (SGS) terrain statistics. Then, these 2-D MWs are assumed to propagate strictly upward instantaneously, as treated by steady 2-D linear theory. The increase in amplitude with decreasing density is taken into account. When the 2-D parameterized wave reaches a large enough amplitude that some instability mechanism (e.g. convective overturning) is predicted to occur, the parameterized wave amplitude is not allowed to grow in amplitude further (e.g. Lindzen 1981) and is said to be "saturated." The reduction in amplitude below that predicted by linear theory results in a reduction of the momentum flux with height, from which the force on the flow is calculated.

Many aspects of current MWD parameterizations are inconsistent with current understanding of MW dynamics and observations. For example, terrain is 3-D and forces 3-D MWs. 3-D MWs can and do propagate laterally, spreading out with height (e.g. Sato et al. 2012; Eckermann et al. 2015). Lateral propagation spreads out MW activity, reduces MW amplitudes, increases breaking heights, and also spreads out the drag relative to current parameterizations, which propagate MWs only vertically. Temporal evolution of the ambient wind and MWs can be important (e.g. Chen et al. 2005; Kruse and Smith 2018) and is also not represented in current parameterizations. Many more deficiencies could be listed.

Additionally, MWD parameterizations are not well constrained by observations. Numerous data sets collected by numerous platforms do contain MW signals; however, these data are often infrequent, are sparse, or incompletely sample the MW field (e.g. data from field campaigns, radiosondes, commercial aircraft). These limitations prevent use of these data to constrain wave properties and the inevitable tuning parameters within MWD parameterizations used by global weather prediction and climate models. Noteworthy possible exceptions are data sets collected by satellite-borne nadir and limb infrared radiometers, which can be sensitive to the temperature perturbations caused by MWs. Recent efforts (e.g. Alexander et al. 2009; Ern et al. 2017; Hindley

et al. 2020) have developed methods of using these satellite data to estimate the momentum flux vectors at stratospheric altitudes, which potentially have the spatial and temporal sampling needed to constrain MWD parameterizations. However, these sensors have limitations as well, being only able to sense a portion of the horizontal and vertical spectrum of scales (e.g. Meyer et al. 2018). The inconsistencies of MWD parameterizations with current physical understanding and their lack of observational constraints motivate additional research.

The overall objective of this work is to provide high-quality estimates of MW momentum fluxes and drags from the troposphere to the mesosphere that are as close to reality as possible. This objective is met via a model inter-comparison involving four state-of-the-science numerical weather prediction (NWP) models that are validated against stratospheric Atmospheric InfraRed Sounder (AIRS) observations. With such a collection of NWP model output, a second route to reality is to evaluate methods of computing MW momentum fluxes from satellite data via method inter-comparison and an observing system simulation experiment (OSSE). For example, momentum fluxes from synthetic AIRS temperature perturbation data, computed as if the AIRS instrument were viewing the modeled atmosphere, can be compared to the true momentum fluxes within the model computed from wind component perturbations. Figure 1 provides an overview of the two approaches to reality. The model validation and inter-comparison is presented here. The method inter-comparison and OSSE are left for a future paper.

The period studied, region of interest, and NWP models used are described in Section 2. The models are quantitatively validated against AIRS data in Section 3. MW properties and spectra are compared between the participating NWP models in Section 4. In Section 5, a previous version and a current MWD parameterization within the Community Atmosphere Model (CAM) are compared with each other and the validated models. The enhancement of zonal drag over the Drake Passage by meridionally-propagating MWs is presented in Section 6. Finally, a summary is provided in Section 7.

2. Case Studied, Participating Models and Their Configurations

a. Region and Period of Interest

The region of focus is centered on the Drake Passage, including the mountains of the Southern Andes, Antarctic Peninsula, and a handful of remote islands in between and to the east (e.g. South

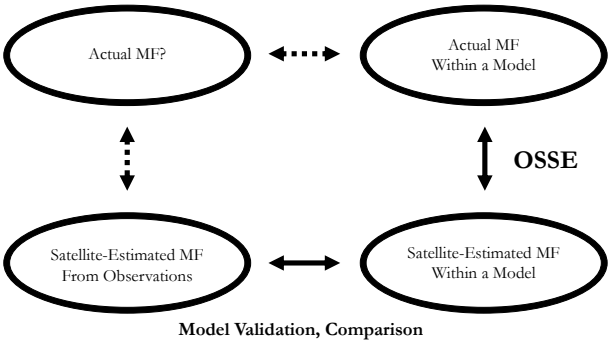


FIG. 1. A graphic showing the types of momentum flux data acquired as part of this project and possible paths (dashed arrows) to estimate actual momentum fluxes from GW observations. For example, momentum fluxes estimated from synthetic AIRS brightness temperature perturbations computed as if the AIRS sensor were viewing through a modeled atmosphere could be compared with the true momentum flux (e.g. $\overline{\rho}\,\overline{u'w'}$) within the model (i.e. an observing system simulation experiment (OSSE)) to understand how satellite-based momentum flux estimates relate to total actual momentum flux. The 2nd route is discussed here, where four GW-resolving models are quantitatively evaluated against AIRS observations.

Georgia). The time period of interest is 9 Oct 2010 through 20 Oct 2010, where many significant MW events were observed by the Atmospheric InfraRed Sounder (AIRS), as shown in Figs. 2 and 3. This period was also chosen as super-pressure balloons flown as part of the CONCORDIASI field campaign transited the Antarctic Peninsula region during this period (Rabier et al. 2010), though, these results are not presented here.

This region is of interest for a number of reasons. The southern Andes mountains are one of the most significant generators of deeply propagating MWs in the world, and certainly the most significant such region in the southern hemisphere (e.g. Hoffmann et al. 2013, 2016; Rapp et al. 2021). The Drake Passage is a region where lateral (i.e. meridional) propagation of these deeply propagating waves into the stratospheric polar night jet (PNJ) is significant (e.g. Sato et al. 2012; Jiang et al. 2013; Hendricks et al. 2014; Amemiya and Sato 2016; Wright et al. 2017; Jiang et al. 2019). In a broader context, this region is of interest as many climate models exhibit the so called "cold-pole problem" (e.g. Eyring et al. 2010), where lower-stratospheric polar temperatures are too cold in the southern hemisphere winter and spring, the meridional temperature gradient is too strong, and (via thermal wind balance) the stratospheric PNJ is also too strong. All of these

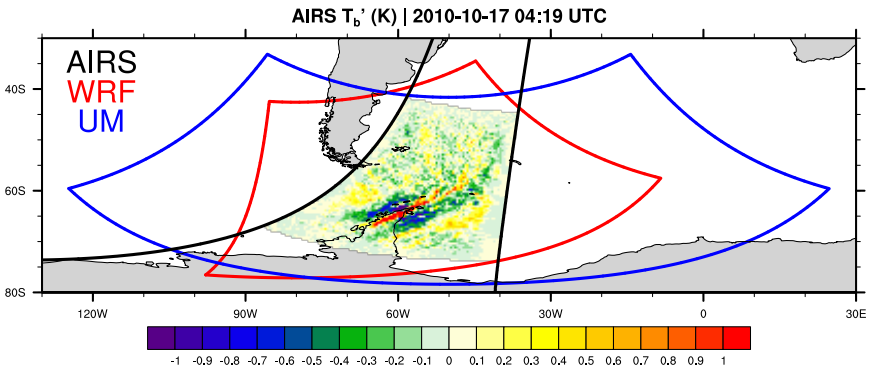


FIG. 2. A map showing the boundaries of the regional $\Delta x = 3$ km (red) WRF and (blue) UM domains. Also shown is an example of what portions of AIRS overpasses were used in the validations of the four models. Here, portions of AIRS swaths not contained within all four model domains were omitted. Additionally, cross-track scans not entirely contained within all domains were also omitted, allowing 2-D Fourier transforms to be used on the rectangular swath data. 4.3 μm AIRS-observed brightness temperature perturbations for an example swath is color shaded where it was used in the model validation. The edges of this swath are shown in black.

errors have important detrimental effects on the simulated climate. McLandress et al. (2012) hypothesized this problem might be caused by missing GW drag in the stratosphere around 60^oS. The sources of these missing GWs and drag in climate models, and whether or not they are responsible for the cold-pole problem, are still debated. Possible sources include missing MWs launched from mountainous islands in this latitude belt (e.g. Alexander and Grimsdell 2013), too little non-orographic GW activity associated with tropospheric jet-stream and frontal imbalances (e.g. Jewtoukoff et al. 2015), and lateral propagation of MWs into the PNJ (e.g. Sato et al. 2012; Amemiya and Sato 2016; Wright et al. 2017).

b. Description of AIRS Observations

Infrared radiances measured by the Atmospheric Infrared Sounder (AIRS) (Aumann et al. 2003; Chahine et al. 2006) on the Aqua satellite are used in this study to evaluate the realism of the stratospheric GW patterns in the simulations. Aqua was launched in May 2002 into a nearly polar, sun-synchronous, low earth orbit at 705-km altitude 100^o inclination and with a 100 min orbital period. Equator crossings occur at 01:30 (descending) and 13:30 (ascending) local times. By scanning across-track, AIRS observes 1780-km wide image swaths with 90 across-track footprints

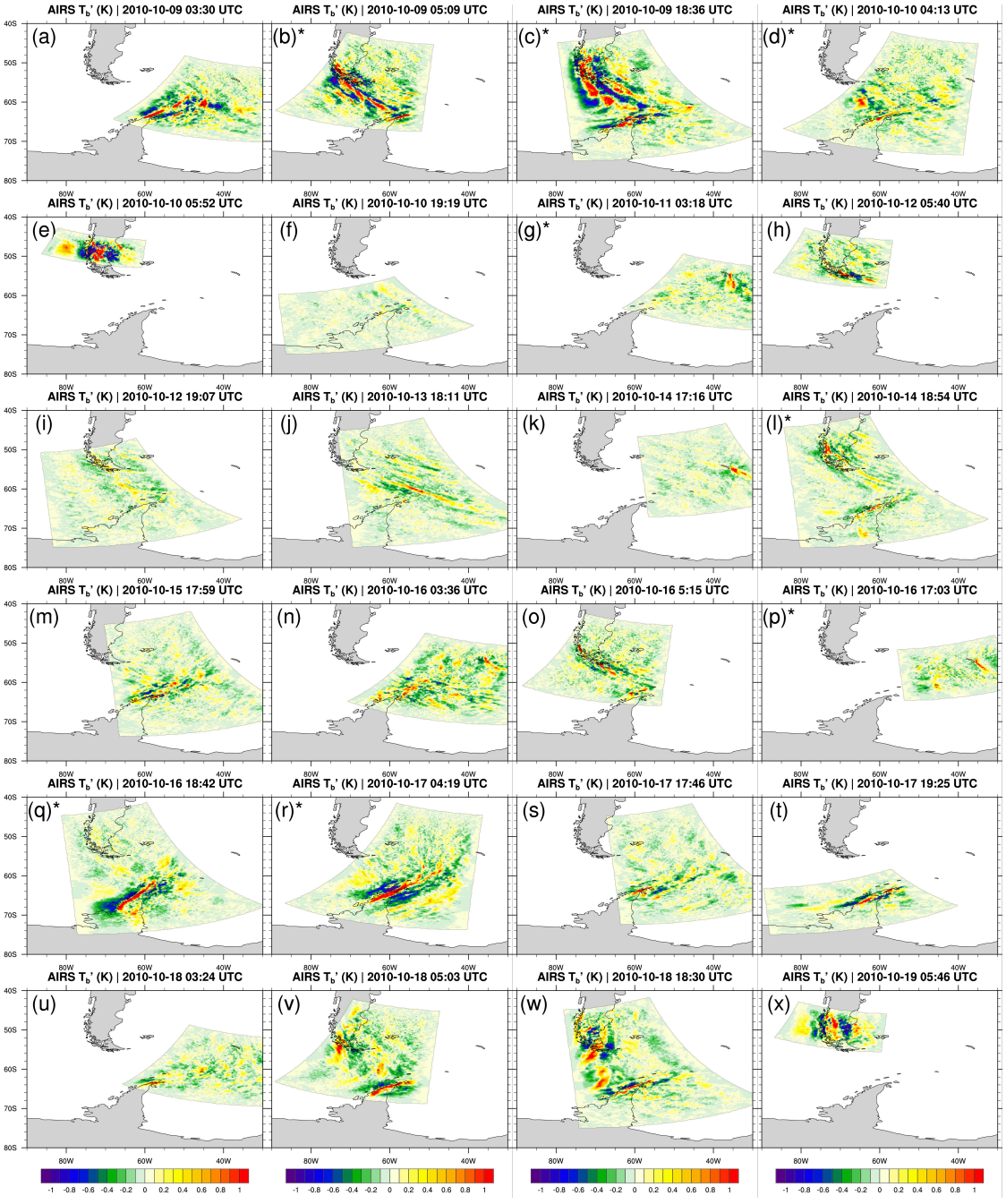


FIG. 3. All AIRS-observed 4.3 μm brightness temperature perturbation swaths used to validate the four participating models (shaded). Areas contoured were contained within all model domains and used to validate the WRF, IFS, and ICON models. Panels with an asterisk were also used in the ICON model validation.

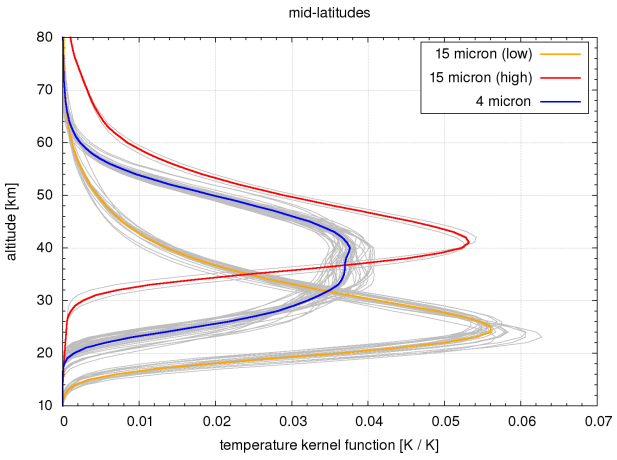


FIG. 4. Individual (thin gray lines) and channel set average (thick colored lines) AIRS instrument weighting functions for midlatitude winter for the (red) 15-μm High product centered at $z \approx 41$ km, the (blue) 4.3-μm product centered at $z \approx 37$ km, and the (yellow) 15-μm Low product centered at $z \approx 24$ km.

and resolution varying from 14 x 14 km^2 at nadir to 21 x 42 km^2 at the swath edge. Adjacent scans are separated by 18-km along-track distance. For observing stratospheric GWs, we average channels over three sets of infrared frequencies sensing the dry stratosphere where clouds, surface emissions, and reflected sunlight do not affect the radiances. Weighting functions for these channel sets (gray lines) and channel averages (colored lines) are shown in Fig. 4. Two of the channel sets lie in the 15 μm band and one in the 4.3 μm band (Alexander and Barnet 2007; Hoffmann et al. 2013, 2017). The vertical width of these weighting functions limit AIRS sensitivity to GWs with vertical wavelengths longer than ≈15 km, with sensitivity increasing for longer vertical wavelengths. Radiances in each channel set are averaged to obtain low-noise brightness temperature data products. Noise varies with scene temperature and channel frequency, but channel-averaged noise varies from 0.1-0.3 K for October at these high southern latitudes (Hoffmann et al. 2014).

GW perturbations are derived as the residual from a cross-track polynomial fit to each scan. Removal of limb-brightening effects is achieved by subtraction of a 4th order polynomial fit to the higher altitude "15 μm (high)" and 4.3 μm sets, and a 6th order fit to the lower altitude "15 μm (low)" set (e.g. Hoffmann et al. 2017). Gravity wave brightness temperature anomalies from AIRS at 4.3 μm are shown in Fig. 3 for all relevant swath segments during the October 9-20, 2010 focus

TABLE 1. A summary of key model configuration parameters. Complete configuration details provided in the text.

Model	$\approx \Delta x$	Domain Top	Sponge Bottom	# Levels	Max Δz
WRF	3	p = 1 Pa (z ≈ 80 km)	z ≈ 70 km	180	520 m
UM	3	z = 85 km	z = 60 km	242	600 m
IFS	9	p = 1 Pa (z ≈ 80 km)	p = 100 Pa (z ≈ 48 km)	137	5 km
ICON	13	z = 80 km	z = 44 km	242	960 m

period. Corresponding swath segments for both 15 μm sets at the same locations are also used in the model validation presented in Section 3.

c. Models and Configurations

Four state-of-the-science NWP models were configured to attempt to reproduce the MWs observed by AIRS over the period of interest (Figs. 2, 3). Two regional models (the Weather Research and Forecasting (WRF) model and the UK Met Office's Unified Model (UM)) and two global models (ECMWF's Integrated Forecast System (IFS) and the German Weather Service's Icosahedral Nonhydrostatic (ICON) model) were used to model the observed waves. Details of individual models' configurations are summarized in Table 1 and described below. With the exception of the IFS, the number and spacing of vertical levels were chosen such that the models had a vertical resolution of at least $\Delta z = 500$ over middle atmosphere below the upper sponge layers. All model initial conditions and boundary conditions were provided by the operational IFS analyses run at $\Delta x \approx 16$ km during this 2010 period.

1) WRF CONFIGURATION

A single $\Delta x \approx 3$ km domain was configured over the region of interest (blue in Fig. 2). The domain was deep, with 180 vertical levels extending up to 1 Pa ($z \approx 80$ km). A 10-km-deep upper sponge layer was used. The vertical resolution varied with height, with the highest resolution near the surface and gradually reduced to ≈ 480 m near the tropopause and kept approximately constant above. Model terrain was derived from the 30 arc-second Global Multi-Resolution Terrain Elevation Data (GMTED) digital elevation model (DEM) (Danielson and Gesch 2011), except for over Antarctica, where the ≈ 200 m resolution Radarsat Antarctic Mapping Project, Version 2

(RAMP2) DEM was used (Liu et al. 2015). The model was initialized once at 12 UTC on 8 Oct 2010 and then integrated for 11 days. During the integration, only boundary conditions from the 6-hourly operational IFS analyses guided the simulated atmosphere. No interior nudging was used within the 3900 km by 3900 km domain. The WRF model (version 4.1.0) required modification to stably integrate beyond a handful of time steps with a domain top this high. For a brief description of these modifications and the physical parameterizations used, see Appendix A.

2) UM Configuration

A global UM simulation, with a grid-spacing of $\Delta x \approx 16$ km and with 85 vertical levels extending to $z = 85$ km initialized every 24 hours from IFS analyses, was used to provide the boundary conditions to a 3000 km by 3000 km regional simulation with a grid-length of $\Delta x \approx 3$ km over the Drake Passage region (red in Fig. 2). Both the global and regional domains were run with UM version 11.1 and used the Global Atmosphere 6 (Walters et al. 2017) and the Regional Atmosphere 1 (Bush et al. 2020) physics configurations, respectively. The $\Delta x \approx 3$ km UM nest was initialized once at 12 UTC on 8 October 2010 and run continuously for 12 days, guided by hourly boundary conditions from the periodically reinitialized global run. The regional UM used a rotated latitude longitude grid, such that the grid-spacing is approximately equal throughout the domain. The vertical resolution was significantly increased within the stratosphere and mesosphere compared with the global simulations, employing 242 vertical levels up to $z = 85$ km where $\Delta z \approx 600$ m.

Similar to the regional WRF simulations, the regional model orography is generated from a blend between the 30 arc-second GMTED DEM and a 1 km version of the RAMP2 DEM. No filtering was performed on the mean orography. A weak implicit Rayleigh damping of the vertical winds (similar to Klemp et al. 2008) is applied from $z \approx 60$ km upward. The parameterized orographic drag from flow blocking and GW drag is turned off, while the turbulent orographic form drag remained switched on.

3) IFS Configuration

The global IFS model version 45r1 (Haiden et al. 2018) was integrated at TCo1279 horizontal resolution (corresponding to 1279 total wavenumbers in the spherical harmonic expansion and a cubic octahedral grid with approximate uniform grid spacing of $\Delta x \approx 9$ km) and reinitialized with the operational analyses of the time every 24 hours at 12UTC over the period of interest and allowed

to run freely for 48h. To minimize spin-up effects, the model output was analyzed for lead times between 24h and 48h. The vertical domain, from the surface to 1 Pa ($z \approx 80$ km) altitude, was resolved with 137 vertical levels. The vertical resolution gradually coarsens with height, such that near the surface the vertical resolution is $\Delta z = 10$ m, in the upper troposphere/lower stratosphere the vertical resolution is $\Delta z \approx 300$ m, and near the model top it is $\Delta z \approx 5$ km. Over the region of interest, the IFS uses a combination of two data sets for model terrain: i) The SRTM30 30 arc-second data set of the Shuttle Radar Topography Mission (Farr et al. 2007) between 60N and 60S, and ii) the RAMP2 data set south of 60S. Note that the resolved terrain in the IFS is represented in spectral space. The smallest resolved wavelength, $2\pi r_e \cos(\phi)/NSMAX$, is sampled by four grid points, where NSMAX is the spherical harmonic truncation (i.e 1279). The IFS here included all physical parameterizations present in version 45r1, including the MW parameterization of Lott and Miller (1997), the turbulent orographic form drag of Beljaars et al. (2004), and the non-orographic GWD parameterization of Scinocca (2003) with the configuration of Orr et al. (2010). To prevent wave reflection from a rigid upper boundary condition, a weak sponge layer in a form of a fourth order hyper-diffusion operator on temperature, vorticity and divergence is applied from 1000 Pa upwards. An additional very strong sponge only on divergence is applied from 100 Pa upwards. It is this very strong divergence-only sponge that is most detrimental for resolved GWs, as will be discussed.

4) ICON Configuration

The ICON (ICOsahedral Non-hydrostatic) modeling framework is a joint project between the German Weather Service and the Max Planck Institute for Meteorology Hamburg for developing a unified next-generation global NWP and climate modeling system (Zängl et al. 2015; Dipankar et al. 2015; Heinze et al. 2017; Giorgetta et al. 2018). The ICON simulations performed here are based on the ICON release version 2.5.0 using the NWP physics parameterizations. To achieve a desired horizontal resolution of 13 km, ICON has been run in resolution R3B7 globally. Initial meteorological conditions were provided by IFS operational analyses. The simulations were performed on 242 vertical model levels from the surface up to 80 km, with layer thickness gradually increasing from 20 m in the lowermost model layer to 960 m at the model top in a height-based terrain-following coordinate system (Leuenberger et al. 2010). A GW-absorbing upper sponge

similar to Klemp et al. (2008) was implemented as well, beginning at $z = 44$ km. Simulations were initialized at 12 UTC every day between 8 to 16 October 2010 and running for 72 hours each. Orographic information has been interpolated to the model grid from the Global Land One-km Base Elevation Project (GLOBE) (Hastings and Dunbar 1999).

d. Quantitative Comparisons between Models and AIRS

Following the approach and methods of Grimsdell et al. (2010); Orr et al. (2015); Hoffmann et al. (2017); Weimer et al. (2021), great care was taken to *quantitatively* compare AIRS observations of stratospheric GWs with those in the models described above. To do so, simulated AIRS brightness temperature perturbations, T'_b, were computed as if the AIRS sensor were viewing through the simulated atmospheres. The Juelich Rapid Spectral Simulation Code (JURASSIC) (Hoffmann and Alexander 2009) was used to sample the 3-D simulated temperature and pressure fields along the line-of-sights between the Aqua satellite and the individual AIRS footprints. Then, these line-of-sight temperature and pressure profiles, along with a representative global mean carbon dioxide concentration of 390 ppm for 2010, were used to compute radiances to the AIRS sensor for individual 15 μm and 4.3 μm AIRS channels, taking into account the radiative transfers for these individual channels/wavelengths. These simulated AIRS data are then averaged and detrended same as the actual AIRS data described in Section 2b, allowing quantitative comparison.

Note that the JURASSIC radiative transfer model does not account for conditions out of local thermodynamic equilibrium (non-LTE conditions). During daytime, carbon dioxide molecules are excited by solar radiation, causing non-LTE conditions in the 4.3 μm wavebands (DeSouza-Machado et al. 2007), which can produce disagreement between the actual and simulated 4.3 μm products. The 15 micron wavebands are typically not affected by non-LTE. Despite the potential discrepancies related to non-LTE conditions, the 4.3 micron channel set is still considered a valuable data source for validation, as this set has significantly lower noise than the 15 μm products (Hoffmann et al. 2014).

3. Model Validation

The four NWP models described above were compared against all AIRS overpasses over the region of interest during the 9-20 Oct 2010 period. Figure 3 shows observed brightness temperature

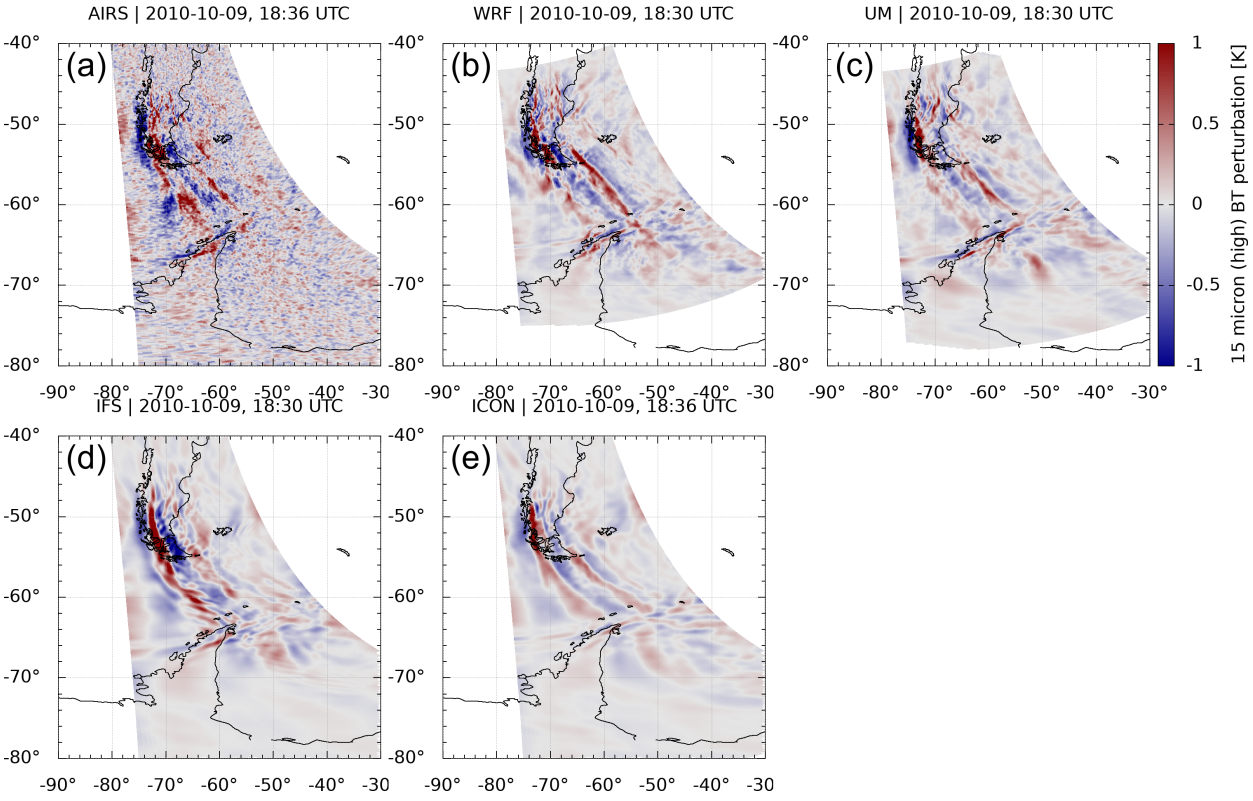


FIG. 5. A qualitative comparison of (a) observed and (b-e) modeled 15-μm-High (z≈41 km) AIRS brightness temperature perturbations for a single AIRS overpass. In this comparison, all models were skillful in reproducing the observed middle-atmosphere MWs. Figures for all overpasses and the three AIRS bands are presented in the supplement to this paper.

perturbations from the 4.3 μm product. Only the data contained within all model domains are shown and included in the model validation below. All 24 overpasses are included in the validation of the IFS, UM, and WRF. Only a selection of these overpasses, indicated by asterisks in Fig. 3, were used in the validation of ICON, due to this model being a later addition to this project.

Figures 5 and 6 show the observed (a) and modeled (b-e) 15-μm-high brightness temperature perturbations for two selected overpasses. While 24 such model-data comparisons were produced for the three collections of AIRS channels, only two examples are shown here: one where all models reproduced the observed middle-atmosphere MWs with remarkable skill (Fig. 5) and one where the models had more moderate skill (Fig. 6). However, similar figures for all overpasses during the period of interest for all three AIRS products are provided in the supplement.

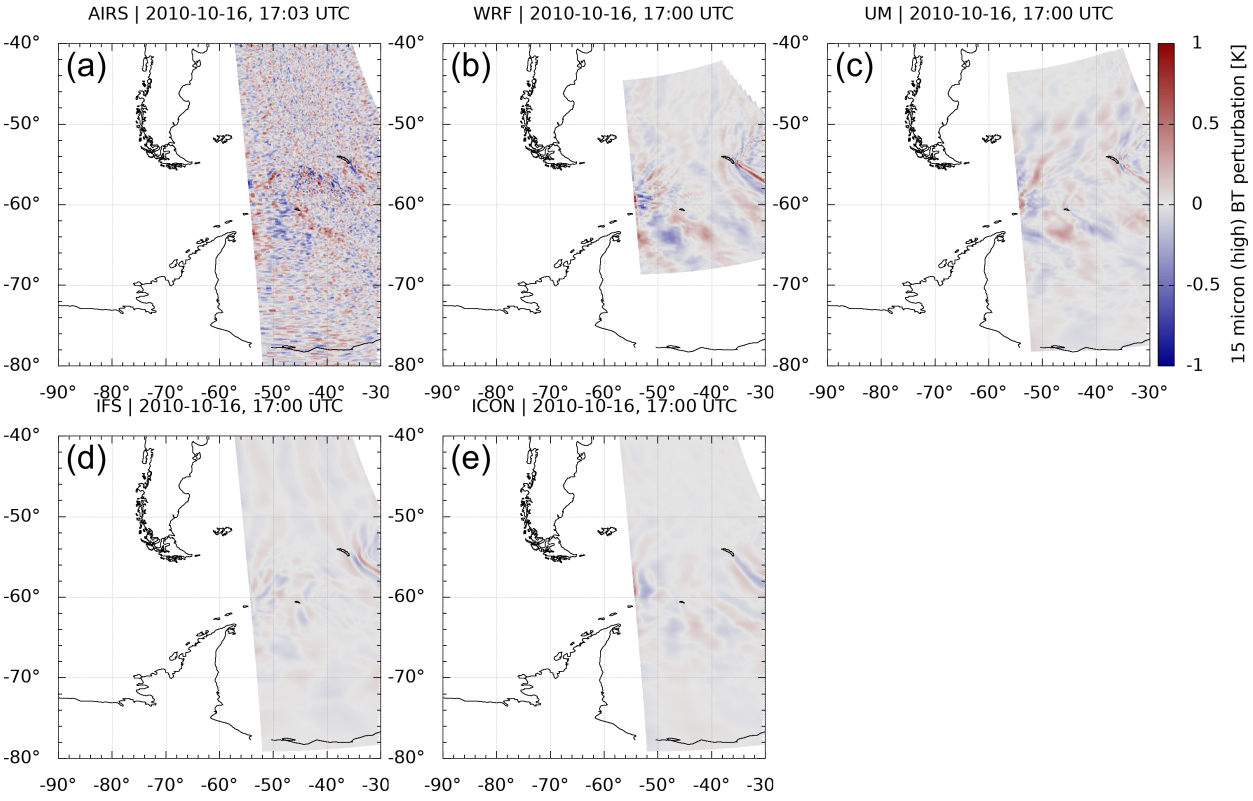


FIG. 6. Same as Fig. 5, but for an AIRS overpass where the models had moderate skill.

Qualitatively, all models had highest skill at reproducing the observed waves from the orographic features containing larger-scale topography (i.e. the Southern Andes and Antarctic Peninsula). The modeled wave structures and amplitudes are reproduced quite well. The models even had some skill in reproducing the observed wave phase. The models had more moderate skill in reproducing waves generated by the smaller-scale terrain of South Georgia (Fig. 6). The waves in the higher-resolution ($\Delta x \approx 3$ km) WRF and UM models have finer scales and overall compare better to the observations than the coarser-resolution IFS ($\Delta x \approx 9$ km) and ICON ($\Delta x \approx 13$ km) models.

Scales and amplitudes of observed and modeled brightness temperature perturbations are compared in Fig. 7, where T_b' spectra are shown. These spectra are computed by first computing the 2-D discrete Fourier transform on each rectangular overpass region shown in Fig. 3. These 2-D spectra were then binned into 1-D wavenumber spectra for each overpass, using the same bin bounds for all overpasses. The variance contributions within each wavenumber bin were then summed over all overpasses to produce the spectra shown in Fig. 7. The observed and modeled T_b' spectra are

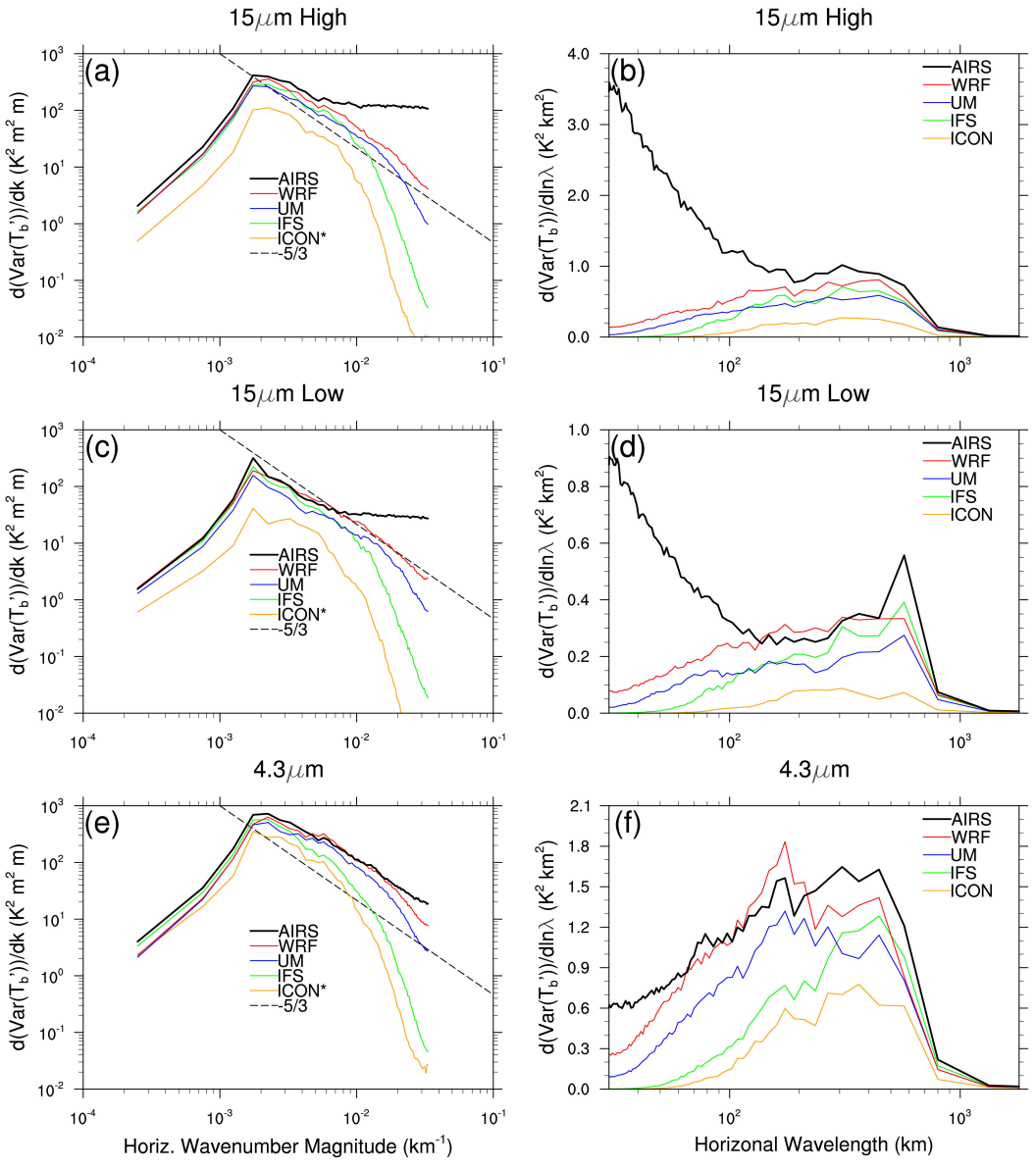
plotted in two ways. Spectra are plotted on a log-log plot (left column) so the slopes of the observed and modeled spectra can be compared to $-5/3$ inertial subrange turbulence spectrum. The spectra are also plotted so the areas under the curves are proportional to the area-integrated T_b' variance ($Var(T_b') = \iint T_b' \, dx \, dy$) on each overpass summed up over all overpasses (right column). Note that the ICON spectra were only summed over a selection of overpasses, resulting in these spectra being offset lower in the plot.

In the log-log spectra (Fig. 7, left column), all three products can be categorized into three wavenumber bands. A drop-off in variance is seen at small wavenumbers (longer scales). This is because only a handful of the overpasses in Fig. 3 contain these longer scales and contribute to these variances. Still, all models were sampled over the same areas shown in Fig. 3, allowing modeled amplitudes at these well-resolved scales to be compared to those observed. Over these longer scales, all models produce amplitudes similar to or slightly less than those observed.

The middle scales are characterized by all spectra roughly having a -5/3 spectral slope. Within this range of scales, the models again produce amplitudes at and just below observations. The amplitudes produced by WRF are the highest and closest to observations, followed by IFS and UM.

Spectra at the smallest scales diverge from each other. While observed amplitudes roughly follow the -5/3 slope in the middle scales, spectra abruptly diverge from this slope, with a more or less flat or "white" spectrum at small scales. This is particularly evident in the 15-μm products, fewer AIRS channels were used to average out noise. This unphysical flat slope at smaller scales suggests instrument noise dominates the observed T_b' variance contributions at these scales. This does *not* suggest that all observations at these scales are noise. Small-scale MWs have been clearly observed in previous studies (e.g. Alexander et al. 2009). However, the spectra presented are representative of the swath-area-integrated $T_b'^2$. While large-amplitude small-scale MWs are definitely present in the observations, smaller-amplitude spatially random noise adds up in the swath-area integral, swamping the small-scale physical features (e.g. South Georgia MWs).

The modeled spectra continue the -5/3 slope further into the smaller scales, but to varying degrees. At the smallest scales resolvable in the simulated AIRS data ($\approx$ 30 km), all models show a notable change in spectral slope. The scale at which this spectral slope changes is roughly interpreted as the effective resolution of each model, showing the scales at which numerical and physical diffusion begin to act on the smallest resolvable scales. These spectra suggest that the



340 FIG. 7. The sum of brightness temperature perturbation spectra for all overpasses plotted on log-log axis (left)
341 and linear-log axis (right). Areas under the curves in (b), (d), and (f) are proportional to area-integrated T_b' in
342 each overpass in Fig. 3 and further summed over all overpasses (i.e. $\propto \sum_{N_{op}} Var(T_b')$, where $Var(T_b') = \iint T_b' \, dx \, dy$
343 and N_{op} is the number of overpasses). Results from the 15 μm High ($z \approx 41$ km, top row), 15 μm Low ($z \approx 24$
344 km, middle row), and 4.3 μm ($z \approx 37$ km, bottom row) products are shown.

369 WRF model may be a bit less diffusive, with a slightly higher effective resolution than the UM

at the same resolution, consistent with the model inter-comparison presented below. The coarser global models have coarser effective resolutions yet, consistent with their coarser grid resolutions.

The contributions to temperature variance by individual bands are best evaluated as plotted in the right column of Fig. 7, where the areas under the curves as plotted are proportional to the sum of the area-integrated $T_b'^2$ over all overpasses. Additionally, the modeled T_b' validation statistics are compared with those observed in Table 2. Note that only scales larger than 100 km and 60 km were used in calculating these statistics for the 15 μm and 4.3 μm products, respectively, in order to exclude scales with white noise in the observations and poor effective model resolutions seen below these scales in Fig. 7. Also note that the statistics in Table 2 were averaged over all available overpasses (24 for WRF, UM, and IFS, 8 for ICON) and so are comparable across all models, while the lower offset of ICON in Fig. 7 cannot be compared with the other spectra. Overall, WRF had the largest amplitudes. Still, WRF under-represented amplitudes at most scales (Fig. 7) and overall in all three products (Table 2). The UM also compared very well with observations, though amplitudes were a bit lower than in WRF, possibly due to having a slightly coarser effective resolution (discussed more below). The coarser IFS underrepresented small-scale MWs, but had amplitudes comparable to WRF and UM at the longer observed scales. ICON produced wave amplitudes similarly close to AIRS as the IFS (Table 2). To summarize, all models were very skillful at reproducing observed middle-atmosphere MWs, allowing quantitative study of middle-atmosphere MWs with some confidence.

TABLE 2. Model error statistics for each AIRS brightness temperature product averaged over all AIRS overpasses available for each model (24 for WRF, UM, IFS and 8 for ICON). $\sigma_{T_b'}$ is the spatial brightness temperature standard deviation for a particular AIRS swath/overpass. $\Delta\sigma_{T_b'}/\sigma_{T_b'}$ is a metric for relative amplitude error of T_b' for a particular overpass. Averaged over all overpasses, this metric is interpreted as a relative T_b' amplitude bias. $|\Delta\sigma_{T_b'}|/\sigma_{T_b'}$ is a relative mean absolute error metric. **NOTE**: All metrics were computed after low-pass filtering T_b' to remove what appears to be white noise in the observed spectra. Scales longer than 100 km (60 km) were retained in the 15 μm products (the 4.3 μm product).

15μm High	WRF	UM	IFS	ICON
$\frac{\Delta\sigma_{T_b'}}{\sigma_{T_b'}}$ (%)	-18.63	-30.41	-32.30	-39.55
$\frac{\lvert\Delta\sigma_{T_b'}\rvert}{\sigma_{T_b'}}$ (%)	19.84	30.41	32.30	39.55
15μm Low	WRF	UM	IFS	ICON
$\frac{\Delta\sigma_{T_b'}}{\sigma_{T_b'}}$ (%)	-8.30	-23.06	-21.34	-29.56
$\frac{\lvert\Delta\sigma_{T_b'}\rvert}{\sigma_{T_b'}}$ (%)	28.70	29.99	22.68	30.28
4.3μm	WRF	UM	IFS	ICON
$\frac{\Delta\sigma_{T_b'}}{\sigma_{T_b'}}$ (%)	-3.94	-18.54	-39.13	-32.75
$\frac{\lvert\Delta\sigma_{T_b'}\rvert}{\sigma_{T_b'}}$ (%)	16.52	19.75	39.45	36.98

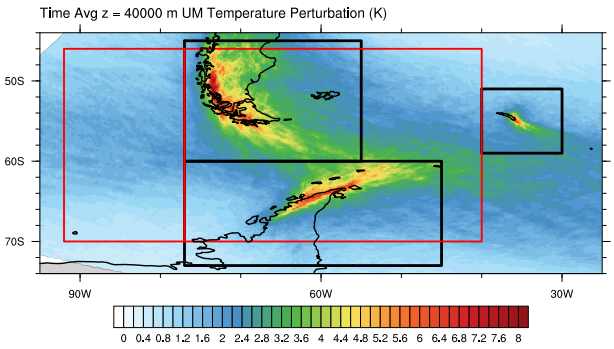


FIG. 8. Map of time-averaged absolute temperature perturbation ($|T'|$) at $z = 40$ km from the UM. Subdomains for the three orographic regions (Southern Andes, Antarctic Peninsula, and South Georgia) used in Figs. 9-15 are outlined in black. The red-outlined regions are those used in the latitude-height sections in Fig. 16.

4. Model Inter-comparison

In the previous section, the models were primarily validated against AIRS observations. Here, the simulations from the different models are compared to one another within the three black-boxed regions in Fig. 8. These subdomains were chosen to be fairly large, so as to mostly contain entire orographic features and the deeply-propagating, laterally spreading waves they generate. Quantities are area-averaged and/or integrated over these subdomains, preventing comparison of spatial distributions of MW activity. Such a comparison is left to future work.

Prior to all analyses here, fields from each model were regridded onto the WRF grid. Perturbation fields were then computed via the spectral filtering method of Kruse and Smith (2015) on the same WRF grid (except for Figs. 8 and 16, where fields were high-pass filtered on the native UM grid to make use of its larger domain), retaining scales smaller than 500 km. Then, fields were further regridded onto regular latitude, longitude grids within each sub-domain, with grids being $\Delta x \approx 3$ km at the northern boundaries. The Southern Andes, Antarctic Peninsula, and South Georgia regions are denoted by "SA", "AP", and "SG". The 10-day average absolute temperature perturbations at $z = 40$ km in the UM simulation are color shaded in Fig. 8, over-viewing the regional MW hot spots and lateral propagation of these MWs in the upper stratosphere.

a. Inter-comparison of MW Forcing

Terrain spectra and slope distributions within each model and sub-domain are compared in Fig. 9. In all sub-domains, all model terrain spectra agree very well at scales larger than 100 km ($k < 10^{-2}$ km^{-1}). All models largely follow the expected -2 slope (Balmino 1993). The notable drop offs of spectra from this slope, and from spectra of the higher-resolution models, give an indication of the effective terrain resolution for each model. The ICON model, which has the coarsest grid resolution ($\Delta x \approx 13$ km), has the coarsest effective terrain resolution as well. The IFS has a slightly higher effective terrain resolution, most easily diagnosed by the precipitous drop off of its terrain spectra. The local maximum at smaller scales in the IFS spectra is spurious. The terrain is represented spectrally, with scales smaller than $4\Delta x \approx 36$ km removed. However, IFS terrain were provided on a TCo1279 grid. The spectral synthesis onto the higher-resolution grid in physical space introduces this small-scale noise in the terrain field. The WRF and UM models have the same grid resolution ($\Delta x \approx 3$ km). However, they clearly have a different effective terrain resolution, with WRF smoothing the grid-mean orography while UM did not.

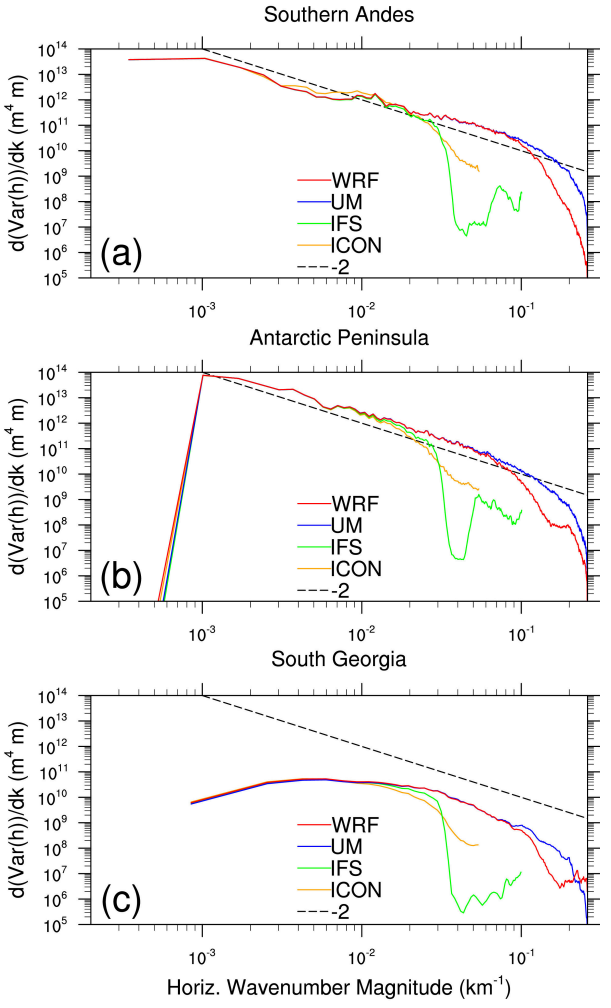
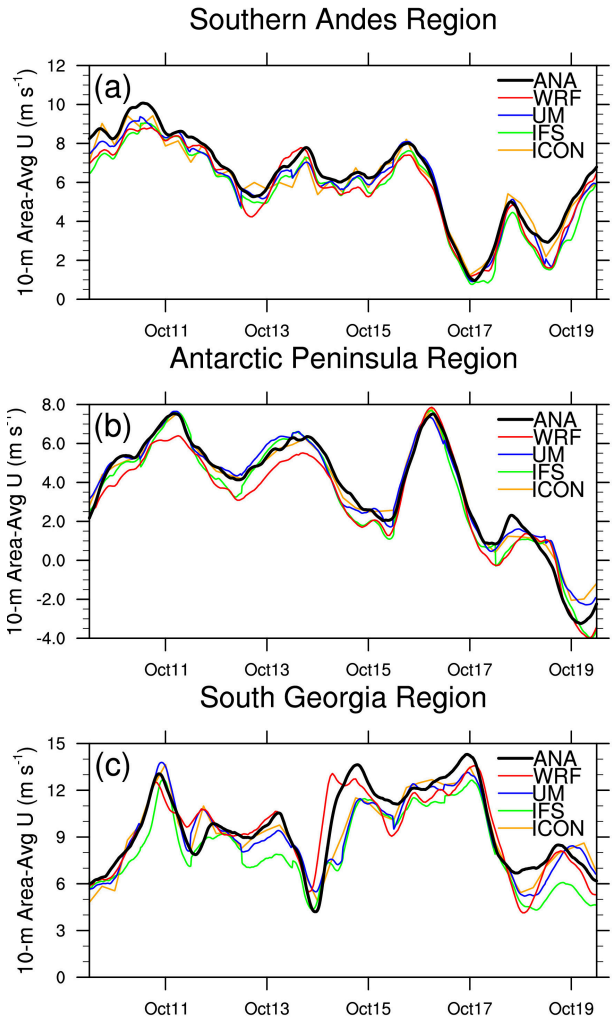


FIG. 9. Model terrain spectra for the four participating models over the three subdomains. Prior to analysis, terrains of all models were regridded onto a common latitude/longitude grid with a resolution of ≈ 3 km. Note that $Var(h) = \iint h^2\, dx\, dy$.



439 FIG. 10. Comparison of subdomain-averaged 10-m zonal winds. The black curve labeled "ANA" shows the
440 subdomain-averaged 10-m winds in the operational IFS analyses used to force all models.

432 The subdomain-averaged 10-m zonal wind time series in each simulation are shown in Fig. 10,
433 along with the subdomain-averaged 10-m winds in the IFS analyses used for initial and/or boundary
434 conditions in all models. These time series of winds that largely force the MWs seen aloft show
435 that despite the different ways each model incorporated IFS analyses (see Section 2c for details) and
436 domain sizes, the models largely agree on these low-level winds and do not drift too far from the
437 IFS analyses, at least over the SA and AP domains. Larger differences are notable in the smaller,
438 more remote SG domain to the east.

b. Overview of Period of Interest

An overview of the 10-day period of interest is provided in Fig. 11, where subdomain-average zonal and meridional winds, vertical fluxes of zonal momentum, and zonal MWD are shown for each subdomain. Only results from the UM model are presented, with this choice being arbitrary. Complete figures for all other models are provided in the supplement.

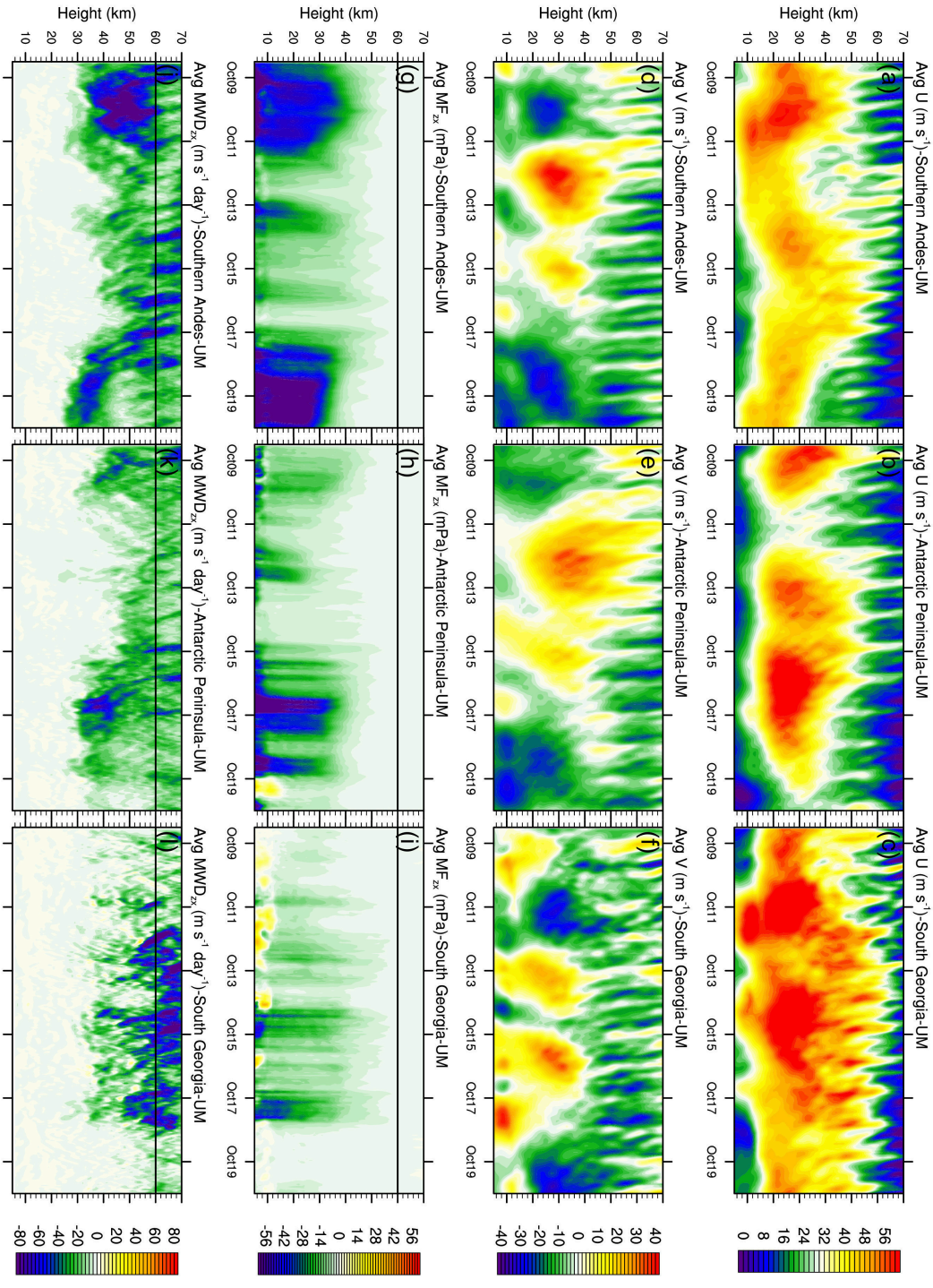


FIG. 11. (top row) Area-averaged zonal wind, (2nd row) meridional wind, (3rd row) MF_{zx}, and (bottom row) GWD_{zx} for the (left column) Southern Andes, (middle column) Antarctic Peninsula, and (right column) South Georgia subdomains. Results shown here are from the Δx = 3 km UM domain. The thin horizontal black line in the bottom two rows indicates the bottom of the 25-km deep sponge used in the UM. A complete model inter-comparison is presented for each subdomain in the supplement to this paper.

Here, the vertical fluxes of horizontal momentum are defined by

$$\begin{aligned} MF_{zx} &= \overline{\rho}\,\overline{u'w'}, \\ MF_{zy} &= \overline{\rho}\,\overline{v'w'}, \end{aligned} \tag{1}$$

and GWDs as

$$\begin{aligned} GWD_{zx} &= -\frac{1}{\overline{\rho}}\frac{\partial MF_{zx}}{\partial z}, \\ GWD_{zy} &= -\frac{1}{\overline{\rho}}\frac{\partial MF_{zy}}{\partial z}. \end{aligned} \tag{2}$$

Because these subdomains are centered on significant orographic features, here, GWD is referred to as MWD.

The PNJ is apparent in the zonal winds within all subdomains, allowing deep vertical propagation of MWs forced by the terrain below. The height and strength of the PNJ varies considerably over the period of interest. Interestingly, strong diurnal variability in both zonal and meridional winds is apparent above the PNJ. This oscillation is presumably some representation of atmospheric tides within the IFS analyses. Whether or not these tides are realistic was not evaluated. However, this large-scale temporal variability was present in all models, strongly influencing the mean environment upon which the MWs propagate and forced diurnal periodicity in MWD_{zx} in the mesosphere (bottom row of Fig. 11). Below these tidally influenced depths, strong MWD_{zx} is evident within the upper-half of the PNJ, where the negative shear forces MW amplitude growth, non-linearity, and breaking, similar to the so-called lower-stratospheric MW valve layer above the subtropical jets (Kruse et al. 2016). Even within this short 10-day period, there is significant temporal variability in vertical fluxes of zonal momentum and drag. MWD_{zx} was quite significant as well, exceeding -50 m s^{-1} day^{-1} frequently and occasionally exceeding -80 m s^{-1} day^{-1} in all subdomains.

c. Model Inter-Comparison of Momentum Flux and Drag

Area-averaged vertical fluxes of horizontal momentum and associated drags are compared between the models in Figs. 12-14. Figs. 12 and 13 show the zonal and meridional components, respectively, as a function of height and time, area-averaged over the South Georgia subdomain. This region was chosen as it had the smallest-scale topography of the three sub-regions, along

with the most significant terrain anisotropy, making it a more difficult test for the models (and parameterizations).

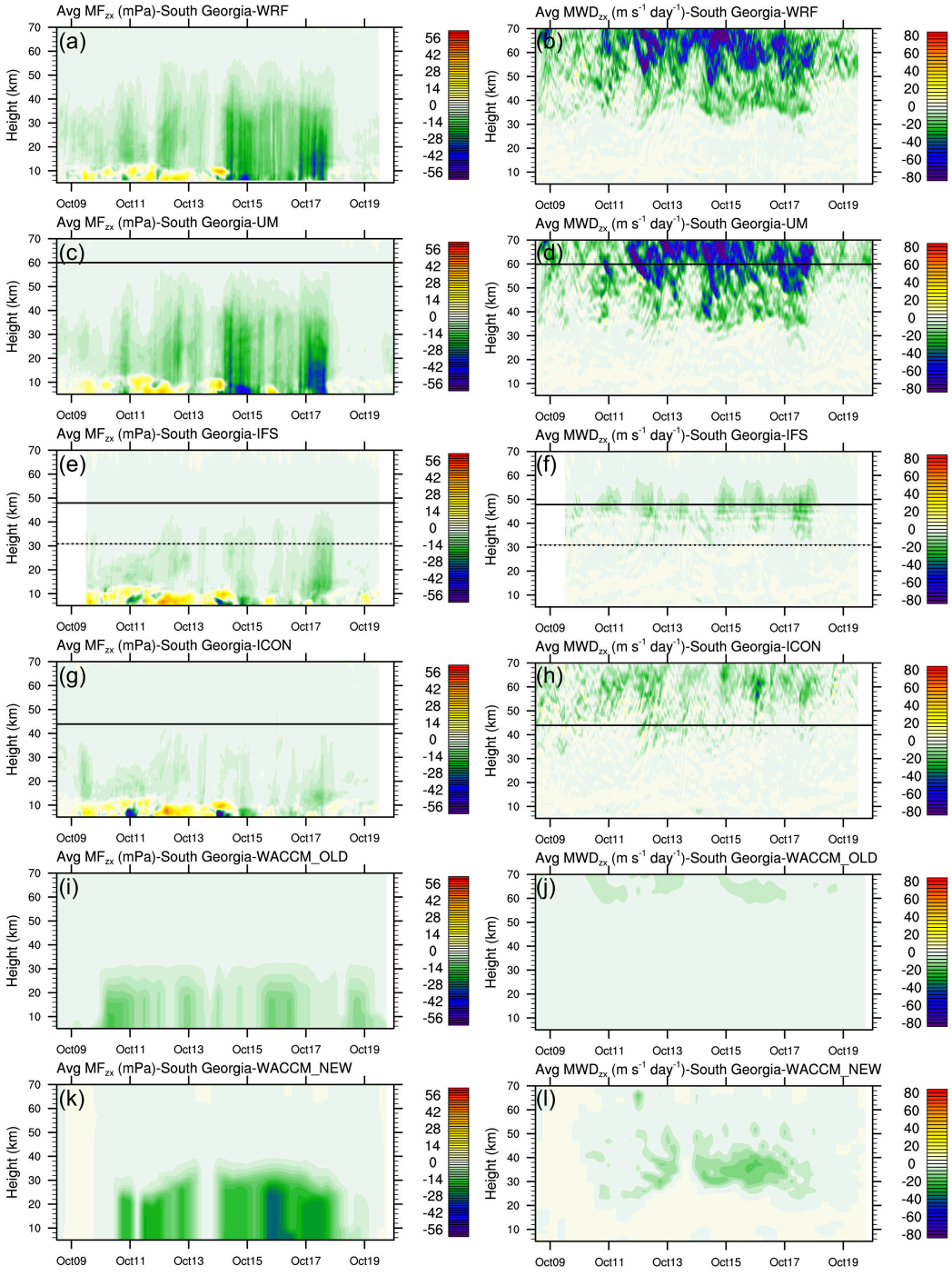


FIG. 12. Comparison of resolved vertical fluxes of zonal momentum and zonal GWD within the four NWP models for the South Georgia subdomain. In the first and second rows, horizontal lines indicate the bottoms of the only sponge layer. In the third row, the horizontal line indicates the bottom of the strong sponge layer in the IFS, and the dashed line indicates the bottom of the first weaker sponge layer. Parameterized fluxes and drags from the previous and current parameterization in CAM are also shown.

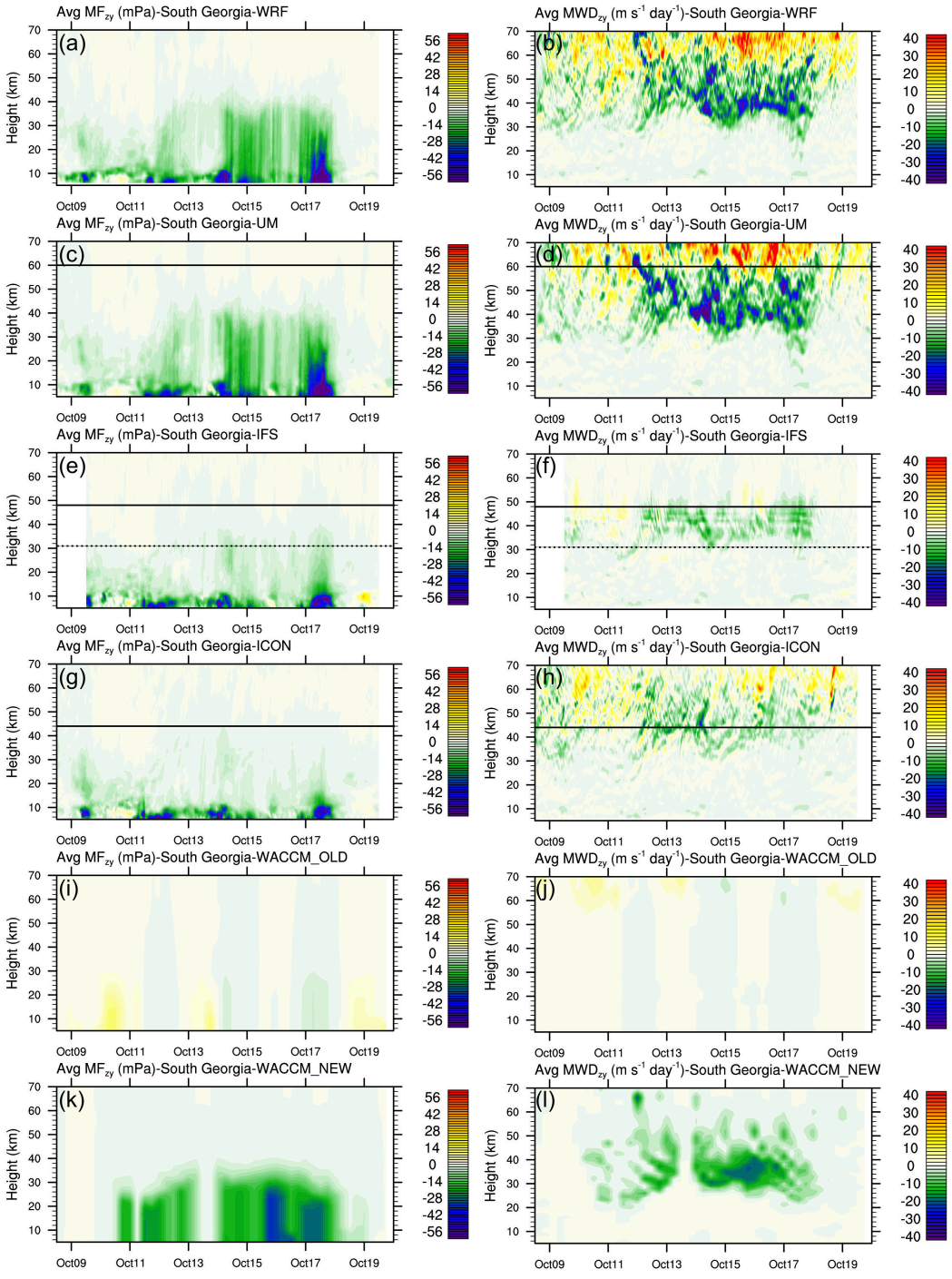
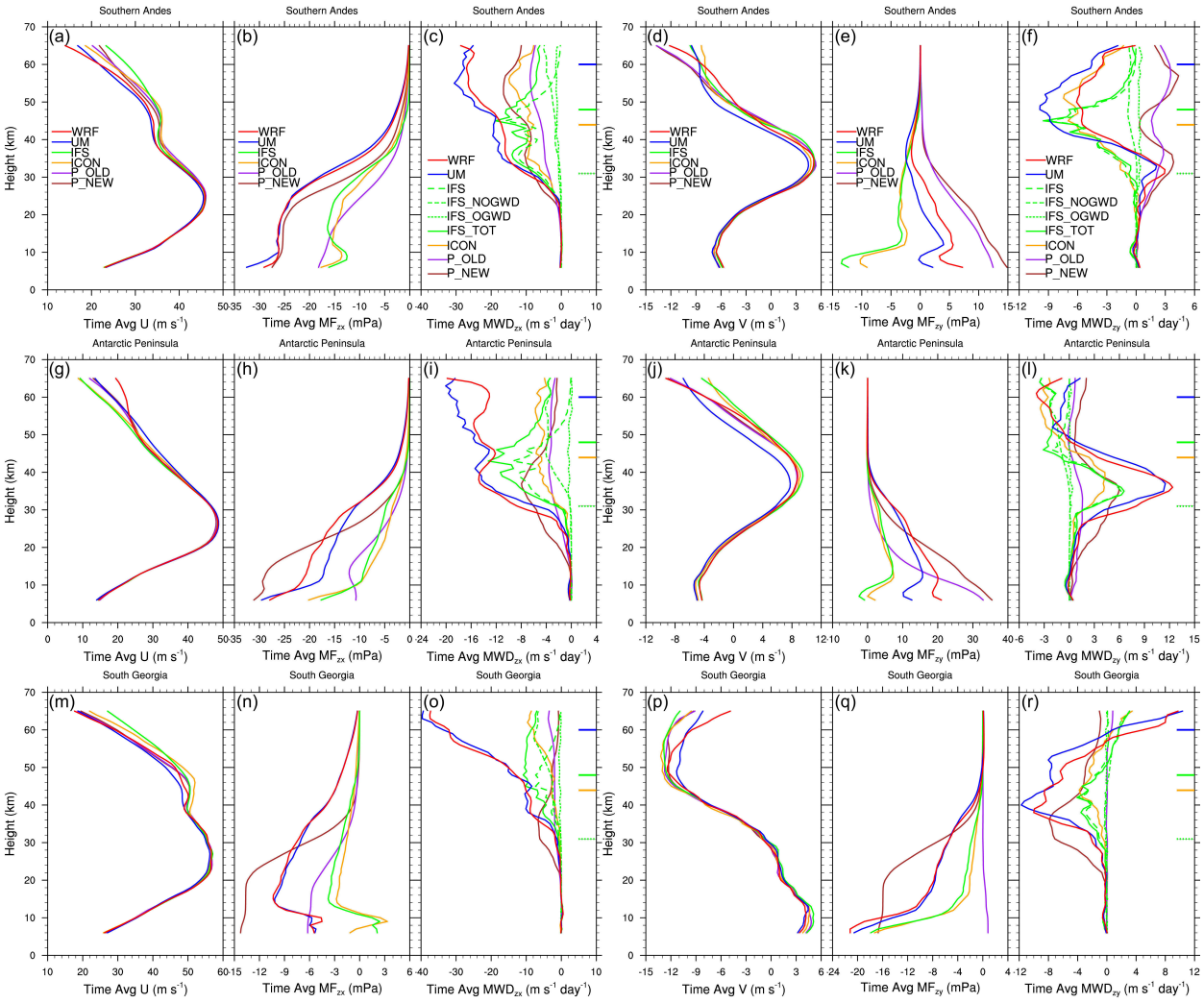


FIG. 13. Same as Fig. 12, but for the meridional components.



481 FIG. 14. Time-averaged profiles of zonal and meridional wind, MF, and MWD averaged over the three
482 sub-domains. Both the resolved and parameterized MFs within the NWP models and CAM parameterizations
483 are shown. In the MWD panels, the WRF, UM, and ICON profiles show the resolved MWDs. MWDs from
484 the previous and current MWD parameterization in CAM are shown as P_OLD and P_NEW, respectively.
485 For the IFS, the resolved (IFS), parameterized non-orographic GW drag (IFS_NOGWD), parameterized MWD
486 (IFS_MWD), and total GWD (IFS_TOT) profiles are shown in green. The bottom of the upper sponge layers
487 are shown via the colored tickmarks on in the drag columns.

The similarly high-horizontal-resolution WRF and UM simulations compared quite well to each other, with quite similar momentum fluxes and drags in both components. The coarser IFS and ICON simulations differ markedly from the $\Delta x \approx 3$ km model runs, having much reduced momentum fluxes and drags in both components and in all subdomains (see supplement). This reduction in wave quantities is likely due to coarser resolution below $z \approx 40$ km and a combination of resolution and the GW-absorbing upper sponges, beginning at $z \approx 40$ km (i.e. 100 Pa) and $z = 44$ km in the IFS and ICON runs, above.

The salient feature in the meridional drag profiles is the the reversal of meridional drag, with southward drag occurring at $z \approx 45$ km, with northward drag further aloft (Figs. 13-14). This may be a result of the strong terrain anisotropy of South Georgia Island, where the predominantly westerly low-level winds impinging on the northwest-to-southeast-oriented terrain would tend to launch southward-propagating MWs with larger amplitudes than those propagating northward. This asymmetry would result in southward drag being deposited at lower altitudes.

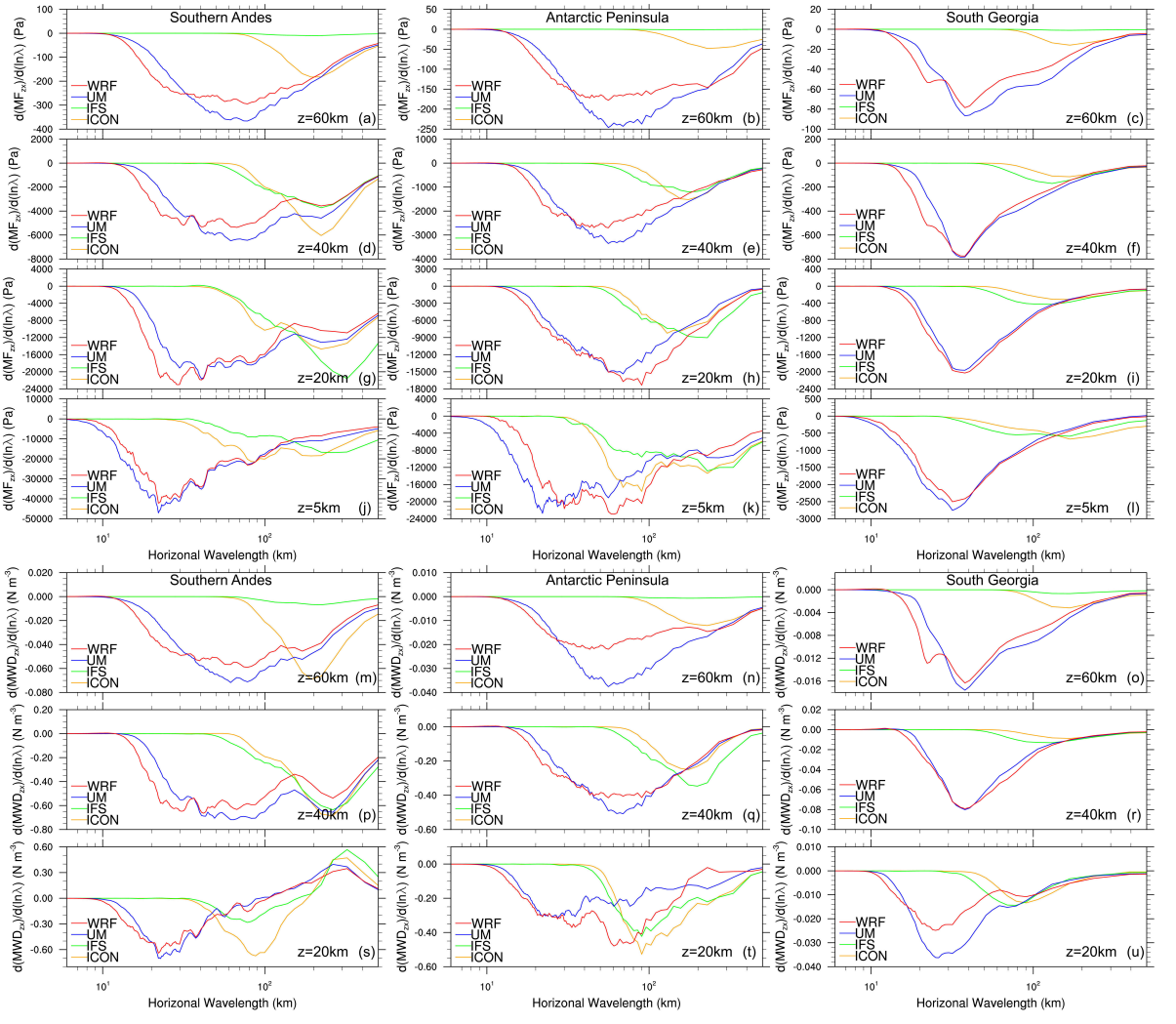
The x-y-time-averaged profiles in Fig. 14 allow more quantitative comparison between the resolved momentum fluxes and drags over the period of interest. Overall, fluxes and drags in the $O(10\,km)$ global models are significantly less than the $\Delta x \approx 3$-km resolution limited-area models. For example, lower-stratospheric resolved zonal momentum fluxes (MF_{zx}) in the global models are about half of those in the higher-resolution limited area models (LAMs) over the SA and AP subdomains, and about a third of those in the SG subdomain. Significant disagreement is apparent in the net meridional momentum fluxes, where different resolutions produce different signs of net flux in the troposphere and lower stratosphere.

Evaluating the effect of resolution on drag is more difficult here, due to the strong GW-damping upper sponges applied in the global models. Despite significantly smaller fluxes into the stratosphere in the global models, the resolved drags are more comparable to the limited-area models at and below $z \approx 45$ km. Further aloft, the upper sponges in the global models have clearly reduced resolved drags, particularly in the IFS output. In the IFS, the orographic and non-orographic GWD parameterizations were active and do compensate for this lack of resolved drag; but still, the total (resolved + parameterized) GWD in the IFS is about a sixth of that in the LAMs in the mesosphere. Sponge or no sponge, MWD parameterizations are still necessary even at the current operational NWP resolutions of $\Delta x \approx 10$ km.

d. Model Inter-Comparison of MW Spectra

The analyses presented in the previous subsection show the net area-integrated momentum fluxes across the entire MW spectrum. MWD parameterizations typically only represent one or two horizontal MW scales in order to predict this area-integrated/averaged flux and drag. However, broad-spectrum orography forces a broad-spectrum of MWs. Here, spatial spectra are presented to inform future MWD parameterization development.

The MF and MWD co-spectra are presented in Fig. 15 at selected altitudes within each subdomain. These spectra were computed by using 2-D FFTs on the subdomain latitude/longitude grids to compute 2-D MF_{zx} co-spectra. These 2-D co-spectra were then binned by wavenumber magnitude into 1-D co-spectra. These 1-D MF co-spectra are plotted such that the area-integral in the spectral space plotted is proportional to the area-averaged fluxes in physical space (e.g. $\frac{1}{A}\int_{-\infty}^{\infty}\frac{\partial MF_{zx}}{\partial \ln\lambda}\,\partial \ln\lambda = \frac{1}{A}\iint \overline{\rho}\,u'w'\,dx\,dy$). The drag co-spectra are computed by taking the vertical derivative of the MF_{zx} co-spectra at ±5 km from the altitudes indicated in the panels (e.g. $\Delta MF_{zx}/\Delta z$, where $\Delta z = 10$ km). These panels show what scales are attenuated and responsible for depositing drag on 10-km depths of atmosphere at the selected levels.



533 FIG. 15. An inter-comparison of time-averaged MF_{zx} and GWD_{zx} co-spectra at selected heights indicated
534 in each panel. These co-spectra are shown within the (left) Southern Andes, (middle) Antarctic Peninsula, and
535 (right) South Georgia subdomains. The GWD_{zx} co-spectra shown are $\Delta MF_{zx}/\Delta z$, where Δz is a 10-km depth
536 centered on the heights indicated in the panels.

The comparison of momentum flux and drag spectra in Fig. 15 highlights the breadth of the flux-carrying MW spectrum and how much of this spectrum is not resolved even at the relatively high $\Delta x \approx 10$-km resolution of the global models. The smaller scales resolved in the LAMs that are under-resolved in the global models (i.e. scales smaller than the $\Delta x_{eff} \approx 60$-km effective resolutions of the global models) contribute significantly to the total fluxes over all altitudes and sub-regions shown. The importance of MWs with similar short horizontal scales was also emphasized by Smith and Kruse (2017), where half of aircraft-observed MW momentum flux was contributed by scales smaller than 60 km on average during the DEEPWAVE field campaign. Curiously, fluxes at larger scales are occasionally larger in the coarser global models, compensating for the lack of flux at smaller scales somewhat. Still, MW parameterizations are necessary at current operational NWP resolutions of $\Delta x \approx 10$ km, as the under- and un-resolved scales may contribute the same amount as that resolved at this resolution.

Another interesting pattern in the MF_{zx} co-spectra is that the UM produces more flux near the surface than WRF at the smallest scales at $z = 5$ km, while this is reversed aloft. We posit that the higher MF_{zx} in the UM at small scales is due to the higher effective terrain resolution (i.e. less model terrain smoothing) in the UM runs (Fig. 9). The MF_{zx} in the UM aloft, along with the smaller amplitudes in the UM noted in the validation above, suggests less physical/numerical diffusion present in the WRF runs and a bit higher effective model resolution.

A final feature in Fig. 15 discussed here is the shift to longer scales with altitude at $\lambda_h < 30$ km in both the WRF model and UM. While the total fluxes are significantly reduced, suggesting significant wave attenuation these depths on average, these scales are also under-resolved as the effective resolution has been found to be $\Delta x_{eff} \approx 7\Delta x = 21$ km for WRF (Skamarock 2004). This shift at small scales is likely unphysical, due to the models' numerical dissipation mechanisms preferentially dissipating the smallest resolvable MWs as they propagate through the middle atmosphere. The systematic differences in MF_{zx} and drag spectra at small scales between UM and WRF and the preferential dissipation at the smallest-resolvable scales known to be under-resolved suggest that even at $\Delta x \approx 3$ km resolution, the entire flux-carrying MW spectrum is still not resolved, consistent the validation results above.

5. CAM MWD Parameterization Evaluation

Figs. 12-14 also include comparisons to two MWD parameterizations implemented within version 6 of the Community Atmosphere Model (CAM, Neale et al. 2010), the general circulation model within the Community Earth System Model (CESM, Danabasoglu et al. 2020). CESM/CAM was used in the WACCM configuration (Marsh et al. 2013), using the finite-volume dynamical core on a latitude/longitude grid with a horizontal resolution of about 1 degree and 70 vertical levels between the surface and $\approx$ 150 km. CAM was initialized in early 2010 and run through much of 2010, initialized with and continuously nudged at all scales and grid points to ERA Interim reanalysis (Dee et al. 2011). By nudging to ERA Interim, CAM's resolved state is kept close to the observed state, and the states of the high-resolution models presented above (e.g. Figs. 14a, g, m, d, j, p, supplemental figures), allowing the parameterized momentum fluxes and drags to be quantitatively compared with those resolved by the models presented above.

The two MWD parameterizations briefly compared here are those implemented in the previous and current versions of CAM, versions 5 and 6, respectively. These parameterizations primarily differ in how the source MWs are specified. With a specified source amplitude, both MWD parameterizations then use conventional linear wave and saturation theory to predict amplitude growth with height and levels of wave breakdown and drag.

The previous parameterization is relatively simple and based on the McFarlane (1987) scheme (for details, see Neale et al. 2010). This parameterization uses an isotropic source formulation, specifying the source MWs and MF to be parallel and opposite to the source-level flow. Source wave amplitudes are specified as twice the standard deviation of sub-grid-scale (SGS) orography, unless this wave amplitude is predicted to overturn, in which case the source amplitude is reduced to the saturation amplitude.

The new parameterization, currently implemented in CAM6, varies significantly in how the SGS terrain is represented and how the source MWs are specified. It fits 2-D ridges to the SGS terrain, finding dominant ridges, along with their widths, lengths, heights, and orientations, in the SGS terrain data. This method allows more realistic, larger ridge heights to be used in estimating source MW amplitudes and allows terrain orientation and anisotropy to be taken into account. Further details are provided in Appendix B.

Parameterized momentum fluxes and drags are quantitatively compared against those resolved in the four high-resolution models presented here in Figs. 12-14. Both the zonal and meridional components of momentum flux and drag are improved, though not perfectly, in the new parameterization. The zonal momentum fluxes and drags are much lower in the old parameterization (cf. Figs. 12 i, k and j, l, respectively). The enhancement in MF_{zx} by the new parameterization is likely due to larger source amplitudes being specified, producing a better comparison to WRF and UM. More striking differences, and improvements, are seen in the meridional components (Fig. 13). The previous isotropic parameterization produces nearly zero meridional fluxes and drags, as the source-level winds are primarily westerly. However, the new parameterization represents the strong northwest-southeast terrain orientation, producing southward momentum fluxes and drags aloft. While these meridional drags compare much better with WRF and UM, the southward drag does appear a little low and the northward drags noted further aloft in WRF and UM are non-existent. Overall, the new parameterization is a significant physical and quantitative improvement over the previous one, but there is still room for improvement.

6. Meridional Propagation of MWs and Drag

A primary motivation for studying the Drake Passage region was the previous work pointing to missing stratospheric GWD near 60^oS (e.g. McLandress et al. 2012) and the density of stratospheric MW hot-spots in this region (e.g. Hoffmann et al. 2013, 2016). The source of this missing drag is not settled, with current hypotheses including an under-representation, or no representation, of MWs and their drag generated by small islands near 60^oS, under-representation of non-orographic GWs and drags near 60^oS, and lateral (i.e. meridional) propagation of MWs into these latitudes of interest.

While the orographic sources in the Drake Passage region are not at 60^oS, MWs laterally-launched by 3-D terrain (e.g. Sato et al. 2012) and meridional refraction of MWs by the deep meridional shear of zonal wind on the flanks of the polar night jet (e.g. Jiang et al. 2013, 2019) both contribute to the presence of GWs over and east of the Drake Passage (e.g. Fig. 8). Transient low-level flow forcing can further spread the meridionally-propagating MWs longitudinally. Still, it is unclear how much zonal momentum is fluxed by these meridionally-propagating waves and whether the resulting zonal drags are significant.

Here, the 3-D fluxes of zonal momentum and the influences of the time- and zonal-mean zonal wind were computed within the UM model. These quantities were computed within the two red boxes in Fig. 8, one upstream and one over and downstream of the Drake Passage. The 3-D zonal MF vector,

$$\mathbf{MF}_x = \langle MF_{xx} = \overline{\rho}\,\overline{u'^2}, MF_{yx} = \overline{\rho}\,\overline{u'v'}, MF_{zx} = \overline{\rho}\,\overline{u'w'}\rangle, \tag{3}$$

the influences on the zonal-mean zonal winds by each flux component,

$$GWD_{xx} = -\frac{1}{\overline{\rho}}\frac{\partial MF_{xx}}{\partial x} \tag{4}$$

$$GWD_{yx} = -\frac{1}{\overline{\rho}}\frac{\partial MF_{yx}}{\partial y} \tag{5}$$

$$GWD_{zx} = -\frac{1}{\overline{\rho}}\frac{\partial MF_{zx}}{\partial z}, \tag{6}$$

and the total influence of the small-scale perturbations on the zonal-mean zonal wind,

$$GWD_{tot} = GWD_{xx} + GWD_{yx} + GWD_{zx}, \tag{7}$$

were computed. The zonal, time-averaged MF_{zx}, GWD_{zx}, GWD_{yx}, and GWD_{tot} are plotted top to bottom in Fig. 16. The GWD_{xx} term was found to be much smaller than GWD_{yx} and GWD_{zx} and is therefore not shown. Vectors of the fluxes of zonal momentum in the meridional plane, $\langle MF_{yx}, MF_{zx}\rangle$, are shown in the bottom row as well. These vectors point opposite to the energy flux vectors in the meridional plane. Energy flux vectors (not shown) point in the direction of GW propagation and group velocity, while the MF flux vectors (Fig. 16, bottom) point to where the GWs came from.

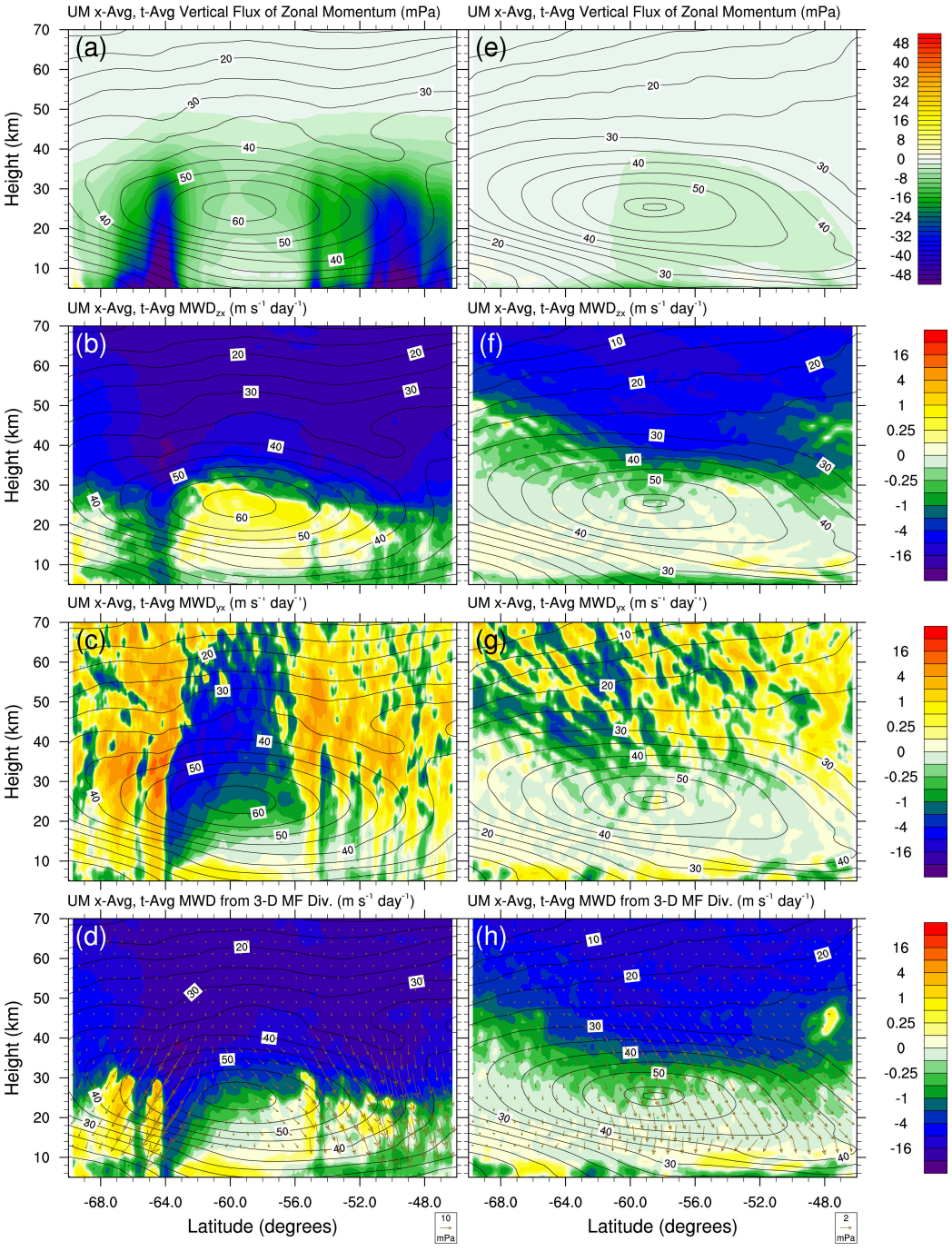


FIG. 16. Latitude-height cross-sections of MF_{zx} (a, e), MWD_{zx} (b, f), MWD_{yx} (c, g), and MWD_{tot} (d, h) within the UM simulation are color shaded. In all panels, zonal wind is contoured. Vectors in (d, h) depict the meridional and vertical fluxes of zonal momentum, with the meridional fluxes scales by the aspect ratio of the plot (1/38). These quantities were averaged over the 10 day simulation and zonally over the regions outlined in red in Fig. 8. The left column shows these quantities over the Drake Passage, while the right column shows these quantities over the Southern Ocean further west.

The MF_{zx} for the Drake Passage region (Fig. 16a) shows lots of flux above the significant terrain features. There is additional MF_{zx} that extends meridionally over the Drake Passage due to meridional propagation of MWs, consistent with the meridional fluxes of zonal momentum (vectors in Fig. 16 d, h) and the clear enhancement compared to further west (Fig. 16e). The fluxes over the Drake Passage near the PNJ maximum are ≈4 times larger. This enhancement suggests that the fluxes and drags over and east of the Drake passage are due to meridional propagation of MWs. The GW fluxes and drags in the box further west are presumed to be primarily from non-orographic sources.

The influence of the vertical divergence of the vertical flux of zonal momentum on the zonal-mean zonal wind is shown in the 2nd row of Fig. 16. The vertically and meridionally propagating MWs produce a local maximum of flux at and above the PNJ maximum, meaning the GWD_{zx} term tends to accelerate the zonal flow at and below the PNJ maximum while exerting the expected drag further aloft. However, the meridional divergence of the meridional flux of zonal momentum term (GWD_{yx}, Fig. 16c) more than compensates the acceleration at and below the PNJ maximum, resulting in a net drag and not acceleration of the PNJ (MWD_{tot}, Fig. 16d). Still, while importance of the meridional divergence of meridional flux of zonal momentum makes physical sense, it should be noted that quadratic terms not involving w' can be sensitive to cutoff length scale chosen to define GW perturbations.

To summarize, meridional propagation of MWs does significantly increase zonal momentum fluxes and drag by a factor of 2-4 over the Drake Passage. The meridional divergence of meridional flux of zonal momentum term was important in the total impact of the small scales on the large-scale zonal flow at and below the PNJ maximum. In fact, if only the vertical divergence of vertical flux of zonal momentum were considered, as is conventionally done, the sign of the drag on the PNJ maximum may be incorrect. These results motivate further research and parameterization development considering lateral propagation of MWs and suggest that consideration of lateral momentum fluxes and resulting drags may be necessary.

7. Summary and Conclusions

In this article, four state-of-the-science NWP models were used to recreate the atmosphere around the Drake Passage during the 10-20 October 2010 period, where numerous MW events

were observed by AIRS within the period. The ICON and IFS models were run at ≈ 13 km and ≈ 9 km resolution in their operational physics configurations. The WRF and UM models were run in a regional configuration at $\Delta x = 3$ km resolution, with vertical resolutions of $\Delta z \leq 600$ m. All models were run with a model top near $z = 80$ km in order to contain the entire life-cycle of the MWs (i.e. from generation to dissipation) and to allow quantitative comparison with AIRS observations. These deep domains allowed most of the AIRS instrument weighting functions for the selected channels to be entirely contained within the domains and be largely unaffected by the necessary upper sponge layers, at least in the WRF and UM domains. Viewing geometries, observation times, and radiative transfers for the AIRS channels/wavelengths used were all taken into account to make this comparison as exact as possible.

The model validation showed that all models had, overall, excellent skill at reproducing the AIRS-observed MWs within the middle atmosphere. This was remarkable, given the numerous differences between the models (e.g. different dynamical cores, parameterizations, grids, initializations, boundary condition methods, source orography data sets) and the deep vertical distances the MWs had to propagate in order to be observed by AIRS. Still, the validations were not perfect. All models appeared to have the most skill reproducing MWs from the Andes, which had the largest-scale, and best-resolved, orography. At the other end of the orographic spectrum, the largest differences between the models and between models and observations occurred for MWs generated by South Georgia Island. Here, it was clear the coarser resolution global NWP models struggled. Despite remarkable qualitative validations, all models under-represented observed MW amplitudes to some extent, even after accounting for effective model resolutions and instrument noise, suggesting even at $\Delta x \approx 3$ km these models still under-resolve and/or over-diffuse MWs.

Resolved momentum fluxes and drags, averaged over the Southern Andes, Antarctic Peninsula, and South Georgia subdomains, were also compared between the four models in detail. The IFS and ICON models, run globally at current operational resolutions of $\Delta x \approx 10$ km, generally had significantly less momentum fluxes and drags than the $\Delta x = 3$ km resolution WRF and UM output. The upper sponges in both the IFS and ICON models had clearly significantly, and artificially, reduced fluxes and drags in the upper stratosphere and mesosphere. While the orographic and non-orographic GWD parameterizations compensated for this reduction somewhat, the total drag was still about a sixth of that resolved by the $\Delta x = 3$ km models.

The inter-comparison of time-averaged zonal momentum flux and drag co-spectra highlighted the breadth of the flux-carrying MW spectrum and how much of the spectrum these $\approx$ 10-km global models are missing relative to the $\approx$ 3-km regional models. These model spectra show the continued need for MW drag parameterizations at current operational NWP resolutions. The systematic differences in MW momentum fluxes at small, grid-scale resolutions between WRF and UM, presumably due to different model (numerical and/or parameterized) diffusions, also suggest the entire flux-carrying MW spectrum is not yet resolved even at $\approx$ 3-km resolution. These results motivate spectral approaches for MW drag parameterizations (e.g. van Niekerk and Vosper 2021).

A significant motivation for this research was to evaluate and ultimately improve MW drag parameterizations. The previous and current MW drag parameterizations in CAM were compared with each other and against the MW-resolving models via CAM runs continuously nudged to the ERA-Interim reanalysis. The current parameterization, which better represents SGS terrain heights and accounts for terrain anisotropy, does seem to be a significant improvement over the previous parameterization, particularly for components of drag perpendicular to the forcing flow. Still, there is clearly lots of room for improvement; there is too much parameterized drag at too low an altitude and too little further aloft.

An interesting feature in the 3-km models was the reversal of meridional MW drag above South Georgia Island. The parameterization did have a representation of the expected southward MW drag, but did not represent resolved drag reversal. If the resolved drag reversal is due to southward-propagating MWs being larger and breaking lower than those coming off the northern slope of South Georgia, then this result further motivates a spectral approach to MW drag parameterization.

Finally, a brief analysis of how meridional propagation of MWs over the Drake Passage influence zonal winds was presented. Meridionally-propagating MWs enhanced the vertical fluxes of zonal momentum over the Drake Passage by about a factor of four. Zonal GWD was enhanced by 200%-400% over the Drake Passage. An interesting result from this analysis was that the meridional divergence of the meridional flux of zonal momentum was important, particularly at and below the PNJ maximum. In fact, here, this term is opposite to, and larger than, the acceleration inferred from the vertical divergence of vertical flux of zonal momentum, which is typically how GWD is defined and computed. Not accounting for this term would result in the wrong direction of drag

being inferred. This result suggests parameterizations that do account for lateral propagation may need to also account for lateral fluxes of horizontal momentum.

Acknowledgments. Numerous institutions and funding agencies supported this international, collaborative work. CGK was supported both by an Advanced Study Program postdoctoral fellowship at NCAR and by the NSF (NSF grant #2004512). MJA was supported by NASA grants #80NSSC18K0069 and #80NSSC17K0169. LH was supported by NASA grants #80NSSC18K0768 and #80NSSC17K0169. PŠ was supported by project CZ.02.2.69/0.0/0.0/19_074/0016231 (International mobility of researchers at Charles University MSCA-IF III). CW was supported by a Royal Society Research fellowship (UF160545). KS was supported by JST, CREST Grant Number JPMJCR1663, Japan. SG was supported by the German Federal Ministry for Education and Research (01LG1907, WASCLIM, ROMIC program). ME was supported by the German Research Foundation (DFG) grant ER 474/4-2 and by the German Federal Ministry of Education and Research (BMBF) grant 01LG1905C (QUBICC, ROMIC). The International Space Science Institute (ISSI) and the Stratosphere-troposphere Processes And their Role in Climate (SPARC) project both supported a meeting and travel for this group. High-performance computing was performed on the Cheyenne supercomputer (ark:/85065/d7wd3xhc) with support provided by NCAR's Computational and Information systems Laboratory, sponsored by the National Science Foundation. The ICON simulations were performed with computing time granted by the John von Neumann Institute for Computing (NIC) and performed on the JURECA supercomputer (Krause and Thörnig 2018) at the Jülich Supercomputing Center (JSC). Additionally, this work used JASMIN, the UK collaborative data analysis facility. Finally, David Gill, Jimy Dudhia, Jordan Powers, Kevin Manning, and Joe Klemp, all within MMM at NCAR, were essential in getting WRF to run in the deep configuration presented here.

Data availability statement. The raw model output on the native grids amounted to more than 20TB, so it is not possible to make all data available for more than a few years. However, post-processed output averaged over the sub-domains used to make the figures will be retained an made available to those who request it. The analysis codes can be made available upon request as well.

APPENDIX A

Deep WRF Configuration

Through many failed attempts, it was learned that WRF, as distributed, cannot be run deeper than ≈ 100 Pa ($z \approx 45$ km) for more than a handful of time steps in the realistic configuration. Three issues were found that had to be overcome.

The first issue was that the default, high-order horizontal interpolators produced intersecting vertical levels after interpolation. David Gill realized that these interpolators are not monotonic, occasionally allowing interpolated values above or below the nearest four analysis points in regions of complex terrain. This caused an intersection of the tightly-spaced vertical levels of the IFS analysis. Using only four point, monotonic horizontal interpolators in WPS prevented this issue.

The second issue was that the vertical interpolation fails at and above ≈ 35 km. This is due to legacy if logic that threw out analysis levels that were too close in pressure. This logic was intended to ignore analysis data on pressure levels close to the surface, assuming the surface analysis to be more accurate. Setting the namelist parameter "zap_close_levels" to 0.1 Pa, from 500 Pa, prevented this problem.

The final issue that prevented stable integrations with a deep domain was instability in the lateral relaxation zone that blends interior, dynamically driven fields with analyses at the lateral boundaries. This instability, often 2 or 3 grid points from the boundary, would cause the solution to blow up in ≈ 20 time steps or less whether or not there was terrain in the domain. Two modifications allowed stable integration with a domain top up to 1 Pa. First, the default lateral relaxation zone was removed by setting "spec_bdy_width" to 1. This namelist modification alone means interior points were 100% dynamically driven, with lateral boundary points primarily having values provided by the analysis. Additionally, WRF's existing grid-point nudging code was modified to replace the default relaxation zone, nudging grid points within 20 grid points of the boundaries to the same analysis used for boundary conditions. This nudging was inversely proportional to distance from the domain boundary. Why these modification worked and the default lateral relaxation did not is an unsolved puzzle. Still, this configuration produced very skillful results, as presented above.

It is not known if modifications to WRF will be made to run released versions in such a deep configuration. However, recent progress has been made in running the Model for Prediction Across Scales (MPAS) in very deep configurations (e.g. Skamarock et al. 2021; Klemp and Skamarock 2021).

The Goddard microphysics scheme (mp_physics = 7), no cumulus scheme, Goddard longwave and shortwave radiation schemes (ra_lw_physics, ra_sw_physics = 5), the Mellor-Yamada-Janjic boundary layer scheme (bl_pbl_physics = 2), the Eta Similarity surface layer scheme (sf_sfclay_physics = 2), and the Unified NOAA land surface model scheme (sf_surface_physics = 2) were used.

APPENDIX B

Overview of Current MWD Parameterization in CAM

The orographic GW (OGW) drag scheme in CAM6 has been modified in two ways from that used in earlier versions of the model (i.e. McFarlane 1987). First, additional drag from flow blocking and high-drag configurations has been incorporated following the approach of Scinocca and McFarlane (2000). Second, the generation of subgrid scale (SGS) forcing data for OGW has been substantially modified from that used in earlier versions of CAM. OGW orientation is now derived from an analysis of ridge orientations (e.g. Baines and Palmer 1990; Lott and Miller 1997; Scinocca and McFarlane 2000). Other OGW parameters are also derived using a different approach.

The new OGW forcing data is derived using a heuristic ridge-finding algorithm (RFA) to identify GW-generating features in the global topography. The RFA will be described in a forthcoming technical note. Here, a brief description is provided to aid in the interpretation of the results presented. Note that the terminology subgrid scale (SGS) was used only for consistency with accepted usage, but in reality what is represented here is unresolved (UnR) topography, which differs from strict SGS topography due to the application of topography smoothing to reduce computational noise (Lauritzen et al. 2015). Henceforth, the term UnR topography is used rather than SGS topography.

Topography processing for CESM begins with global topography binned onto a 3km cubed-sphere grid (Lauritzen et al. 2015), denoted by $H3(x, y)$. The first step in identifying UnR wave generating features is to smooth $H3(x, y)$ with a prescribed smoothing radius that depends on resolution and dynamical core. The smooth topography, $H_s(x, y)$, is used as the bottom boundary for CAM's dynamical core, after conservative remapping to the eventual computational grid. The smoothing radius used to generate $H_s(x, y)$ is determined by trial and error for each dynamical core.

In general, a smoothing radius of 2 to 3 times the horizontal grid spacing used by the dynamical core is required to reduce computational noise.

$H_s(x,y)$ is subtracted from raw 3km topography, $H3(x,y)$, to give the unresolved topography $H'(x,y)$. The RFA tests 16 different orientations, $\phi_k,\ k = 1,2,\ldots,16$, at the locus of each maximum in $H'(x,y)$. Mean ridge elevation profiles are calculated by averaging horizontally along the putative ridge crest. The "best" ridge-orientation is defined as that where the variance of the 16 mean elevation profiles is a maximum. The list of features is thinned by combining adjacent features with approximately the same orientation. Additional relevant GW parameters are calculated for each feature. An obstacle height, h_r, is defined as the difference between the maximum and minimum elevations in the ridge profile. In addition, two length scales for each feature are estimated; a ridge half-width, w_r, and an along-ridge length, l_r. A visual depiction of the process is shown in Fig. B1. Note that these RFA results depend only on the smoothing radius chosen and are completely independent of the eventual model computational grid.

The final step in the generation of the UnR topographic forcing data is to map the ridge-skeleton shown in Fig. B1 from the 3-km starting grid onto the desired atmosphere model computational grid. This is done by summing or averaging ridge properties over the 16 original test orientations within each grid box. If a ridge feature on the 3-km grid straddles multiple computational cell areas, its crest length l_r is correctly partitioned across the grid areas. If multiple features with the same orientation are found in a single grid box their lengths are summed and other properties are determined from weighted averages. The final result of the RFA is a gridded set of ridges at 16 orientations with properties $h^*(\phi_k)$, $l^*(\phi_k)$, and $w^*(\phi_k)$.

The gridded forcing parameters from the RFA are then coupled to the orographic gravity wave (OGW) scheme. In contrast to the earlier OGW scheme in CAM (McFarlane 1987), wave orientations are set by the UnR ridge orientations, ϕ_k, and not the low-level wind. In addition, OGW efficiency, which was previously a specified global parameter, is now estimated from topographic length scales in each model grid-box according to:

$$\epsilon(\phi_k) = \alpha_e \frac{w^*(\phi_k) l^*(\phi_k)}{A_g} \tag{B1}$$

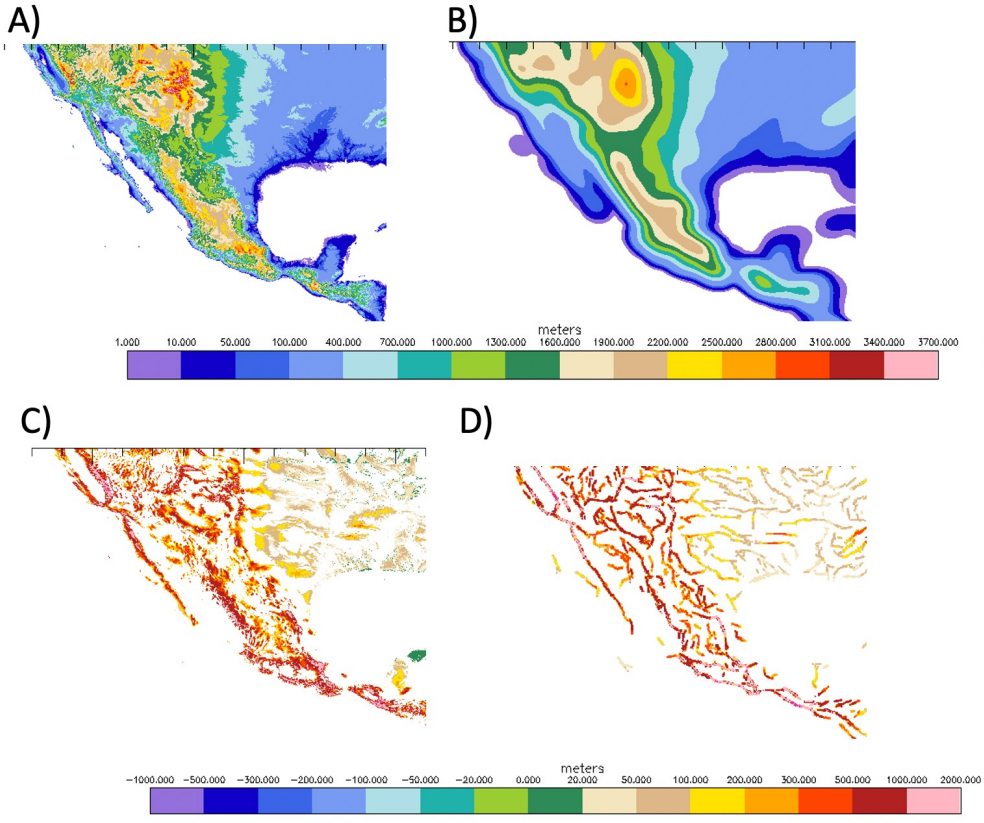


FIG. B1. Steps in ridge-finding algorithm (RFA): (a) 3km binned topography $H3(x,y)$; (b) Smoothed topography $H_s(x,y)$. Note, here a smoothing radius of 180 km has been applied, which is used for 1^o resolution in the CAM latitude-longitude finite-volume dynamical core; (c) Unresolved topography consistent with $H3$ and H_s in (a,b). Negative values are not plotted in order to facilitate comparsion with ridge-skeleton in (d); and (d) Ridge "skeleton" reconstructed on 3-km grid from features identified by RFA. Line segments follow ridges crests. Colors show RFA's estimate of obstacle heights.

where α_e is a tunable parameter and A_g is the grid-cell area. Initial vertical wave displacements are set to $h^*(\phi_k)$. The new OGW scheme calculates enhanced near surface drag from low-level blocking, and high-drag configurations following Scinocca and McFarlane (2000).

This shows how to enter the commands for making a bibliography using BibTeX. It uses references.bib and the ametsocV6.bst file for the style.

References

Alexander, M. J., and C. Barnet, 2007: Using satellite observations to constrain parameterizations of gravity wave effects for global models. *J. Atmos. Sci.*, **64 (5)**, 1652 – 1665, https://doi.org/10.1175/JAS3897.1.

Alexander, M. J., S. D. Eckermann, D. Broutman, and J. Ma, 2009: Momentum flux estimates for South Georgia island mountain waves in the stratosphere observed via satellite. *Geophys. Res. Lett.*, **36**, https://doi.org/10.1029/2009GL038587.

Alexander, M. J., and A. W. Grimsdell, 2013: Seasonal cycle of orographic gravity wave occurrence above small islands in the southern hemisphere: Implications for effects on the general circulation. *J. Geophys. Res.*, **118**, https://doi.org/10.1002/2013JD020526.

Amemiya, A., and K. Sato, 2016: A new gravity wave parameterization including three-dimensional propagation. *J. Meteor. Soc. Japan*, **94 (3)**, 237–256, https://doi.org/10.2151/jmsj.2016-013.

Aumann, H., and Coauthors, 2003: Airs/amsu/hsb on the aqua mission: design, science objectives, data products, and processing systems. *IEEE Transactions on Geoscience and Remote Sensing*, **41 (2)**, 253–264, https://doi.org/10.1109/TGRS.2002.808356.

Baines, P. G., and T. N. Palmer, 1990: Rationale for a new physically based parameterization of subgrid scale orographic effects. Tech. rep., ECMWF, 11 pp. pp.

Balmino, G., 1993: The spectra of the topography of the Earth, Venus and Mars. *Geophys. Res. Lett.*, **20 (11)**, 1063–1066, https://doi.org/https://doi.org/10.1029/93GL01214.

Beljaars, A. C. M., A. R. Brown, and N. Wood, 2004: A new parametrization of turbulent orographic form drag. *Quart. J. Roy. Meteor. Soc.*, **130 (599)**, 1327–1347, https://doi.org/https://doi.org/10.1256/qj.03.73.

Bush, M., and Coauthors, 2020: The first Met Office Unified Model–JULES Regional Atmosphere and Land configuration, RAL1. *Geoscientific Model Development*, **13 (4)**, 1999–2029, https://doi.org/10.5194/gmd-13-1999-2020, URL https://gmd.copernicus.org/articles/13/1999/2020/.

Chahine, M. T., and Coauthors, 2006: Airs: Improving weather forecasting and providing new data on greenhouse gases. *Bull. Amer. Meteor. Soc.*, **87 (7)**, 911 – 926, https://doi.org/10.1175/BAMS-87-7-911.

Chen, C.-C., D. R. Durran, and G. J. Hakim, 2005: Mountain-wave momentum flux in an evolving synoptic-scale flow. *J. Atmos. Sci.*, **62 (9)**, 3213–3231, https://doi.org/10.1175/JAS3543.1.

Danabasoglu, G., and Coauthors, 2020: The community earth system model version 2 (cesm2). *J. Adv. in Modeling Earth Systems*, **12 (2)**, https://doi.org/https://doi.org/10.1029/2019MS001916.

Danielson, J. J., and D. B. Gesch, 2011: Global multi-resolution terrain elevation data 2010 (GMTED2010). Tech. rep., U.S. Geological Survey Open-File Report 2011–1073, 26 p.

Dee, D. P., and Coauthors, 2011: The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Quart. J. Roy. Meteor. Soc.*, **137 (656)**, 553–597, https://doi.org/https://doi.org/10.1002/qj.828.

DeSouza-Machado, S. G., L. L. Strow, S. E. Hannon, H. E. Motteler, M. Lopez-Puertas, B. Funke, and D. P. Edwards, 2007: Fast forward radiative transfer modeling of 4.3 μm nonlocal thermodynamic equilibrium effects for infrared temperature sounders. *Geophys. Res. Lett.*, **34 (1)**, https://doi.org/https://doi.org/10.1029/2006GL026684.

Dipankar, A., B. Stevens, R. Heinze, C. Moseley, G. Zängl, M. Giorgetta, and S. Brdar, 2015: Large eddy simulation using the general circulation model ICON. *J. Adv. in Modeling Earth Systems*, **7 (3)**, 963–986, https://doi.org/https://doi.org/10.1002/2015MS000431.

Eckermann, S. D., J. Ma, and D. Broutman, 2015: Effects of horizontal geometrical spreading on the parameterization of orographic gravity wave drag. Part I: Numerical transform solutions. *J. Atmos. Sci.*, **72 (6)**, 2330–2347, https://doi.org/10.1175/JAS-D-14-0147.1.

Ern, M., L. Hoffmann, and P. Preusse, 2017: Directional gravity wave momentum fluxes in the stratosphere derived from high-resolution AIRS temperature data. *Geophys. Res. Lett.*, **44 (1)**, 475–485, https://doi.org/10.1002/2016GL072007.

Eyring, V., T. G. Shepherd, and D. W. Waugh, 2010: SPARC CCMVal Report on the Evaluation of Chemistry-Climate Models. Tech. rep., SPARC, 426 pp. pp. URL http://www.sparc-climate.org/publications/sparc-reports/.

909 Farr, T. G., and Coauthors, 2007: The shuttle radar topography mission. *Rev. Geophys.*, **45 (2)**,
910 https://doi.org/https://doi.org/10.1029/2005RG000183.

911 Fritts, D. C., and M. J. Alexander, 2003: Gravity wave dynamics and effects in the middle
912 atmosphere. *Rev. Geophys.*, **41**, 1–64.

913 Fritts, D. C., and Coauthors, 2016: The deep propagating gravity wave experiment (DEEPWAVE):
914 An airborne and ground-based exploration of gravity wave propagation and effects from their
915 sources throughout the lower and middle atmosphere. *Bull. Amer. Meteor. Soc.*, **97 (3)**, 425–453,
916 https://doi.org/10.1175/BAMS-D-14-00269.1.

917 Giorgetta, M. A., and Coauthors, 2018: ICON-A, the Atmosphere Component of the ICON Earth
918 System Model: I. Model Description. *J. Adv. in Modeling Earth Systems*, **10 (7)**, 1613–1637,
919 https://doi.org/https://doi.org/10.1029/2017MS001242.

920 Grimsdell, A. W., M. J. Alexander, P. T. May, and L. Hoffmann, 2010: Model study of waves
921 generated by convection with direct validation via satellite. *J. Atmos. Sci.*, **67 (5)**, 1617 – 1631,
922 https://doi.org/10.1175/2009JAS3197.1.

923 Haiden, T., M. Janousek, J.-R. Bidlot, R. Buizza, L. Ferranti, F. Prates, and F. Vitart, 2018: Evalu-
924 ation of ecmwf forecasts, including the 2018 upgrade. Tech. Rep. 831, ECMWF. https://doi.org/
925 10.21957/ldw15ckqi, URL https://www.ecmwf.int/node/18746.

926 Hastings, D. A., and P. K. Dunbar, 1999: Global Land One-kilometer Base Elevation (GLOBE).
927 Tech. rep., U.S. Dept. of Commerce, National Oceanic and Atmospheric Administration, Na-
928 tional Environmental Satellite, Data, and Information Service. URL http://www.ngdc.noaa.gov/
929 mgg/topo/globe.html.

930 Heinze, R., and Coauthors, 2017: Large-eddy simulations over Germany using ICON: a
931 comprehensive evaluation. *Quart. J. Roy. Meteor. Soc.*, **143 (702)**, 69–100, https://doi.org/
932 https://doi.org/10.1002/qj.2947.

933 Hendricks, E. A., J. D. Doyle, S. D. Eckermann, Q. Jiang, and P. A. Reinecke, 2014:
934 What Is the Source of the Stratospheric Gravity Wave Belt in Austral Winter? *Journal
935 of the Atmospheric Sciences*, **71 (5)**, 1583–1592, https://doi.org/10.1175/JAS-D-13-0332.1,

URL https://doi.org/10.1175/JAS-D-13-0332.1, https://journals.ametsoc.org/jas/article-pdf/71/5/1583/3647150/jas-d-13-0332_1.pdf.

Hindley, N. P., C. J. Wright, L. Hoffmann, T. Moffat-Griffin, and N. J. Mitchell, 2020: An 18-year climatology of directional stratospheric gravity wave momentum flux from 3-d satellite observations. *Geophys. Res. Lett.*, **47 (22)**, https://doi.org/https://doi.org/10.1029/2020GL089557.

Hoffmann, L., and M. J. Alexander, 2009: Retrieval of stratospheric temperatures from atmospheric infrared sounder radiance measurements for gravity wave studies. *J. Geophys. Res.: Atmospheres*, **114 (D7)**, https://doi.org/https://doi.org/10.1029/2008JD011241.

Hoffmann, L., M. J. Alexander, C. Clerbaux, A. W. Grimsdell, C. I. Meyer, T. Rößler, and B. Tournier, 2014: Intercomparison of stratospheric gravity wave observations with AIRS and IASI. *Atmospheric Measurement Techniques*, **7 (12)**, 4517–4537, https://doi.org/10.5194/amt-7-4517-2014.

Hoffmann, L., A. W. Grimsdell, and M. J. Alexander, 2016: Stratospheric gravity waves at Southern Hemisphere orographic hotspots: 2003–2014 AIRS/Aqua observations. *Atmospheric Chemistry and Physics*, **16 (14)**, 9381–9397, https://doi.org/10.5194/acp-16-9381-2016.

Hoffmann, L., R. Spang, A. Orr, M. J. Alexander, L. A. Holt, and O. Stein, 2017: A decadal satellite record of gravity wave activity in the lower stratosphere to study polar stratospheric cloud formation. *Atmos. Chem. Phys.*, **17 (4)**, 2901–2920, https://doi.org/10.5194/acp-17-2901-2017.

Hoffmann, L., X. Xue, and M. J. Alexander, 2013: A global view of stratospheric gravity wave hotspots located with atmospheric infrared sounder observations. *J. Geophys. Res.: Atmospheres*, **118 (2)**, 416–434, https://doi.org/https://doi.org/10.1029/2012JD018658.

Holton, J. R., 1983: The influence of gravity wave breaking on the general circulation of the middle atmosphere. *J. Atmos. Sci.*, **40**, 2497–2507, https://doi.org/10.1175/1520-0469(1983)040<2497:TIOGWB>2.0.CO;2.

Jewtoukoff, V., A. Hertzog, R. Plougonven, A. de la Camara, and F. Lott, 2015: Comparison of gravity waves in the southern hemisphere derived from balloon observations and the ECMWF analyses. *J. Atmos. Sci.*, **72 (9)**, 3449 – 3468, https://doi.org/10.1175/JAS-D-14-0324.1.

963 Jiang, Q., J. D. Doyle, S. D. Eckermann, and B. P. Williams, 2019: Stratospheric Trail-
964 ing Gravity Waves from New Zealand. *J. Atmos. Sci.*, **76 (6)**, 1565–1586, https://doi.org/
965 10.1175/JAS-D-18-0290.1.

966 Jiang, Q., J. D. Doyle, A. Reinecke, R. B. Smith, and S. D. Eckermann, 2013: A modeling study of
967 stratospheric waves over the Southern Andes and Drake Passage. *J. Atmos. Sci.*, **70**, 1668–1689.

968 Klemp, J. B., J. Dudhia, and A. D. Hassiotis, 2008: An upper gravity-wave absorbing layer for
969 NWP applications. *Mon. Wea. Rev.*, **136**, 3987–4004.

970 Klemp, J. B., and W. C. Skamarock, 2021: Adapting the MPAS Dynamical Core for Applications
971 Extending Into the Thermosphere. *J. of Adv. in Modeling Earth Systems*, **13 (9)**, https://doi.org/
972 https://doi.org/10.1029/2021MS002499.

973 Krause, D., and P. Thörnig, 2018: JURECA: Modular supercomputer at Jülich Supercomputing
974 Centre. *J. of Large-Scale Research Facilities*, **4 (A132)**, https://doi.org/http://dx.doi.org/10.
975 17815/jlsrf-4-121-1.

976 Kruse, C. G., 2018: Mountain wave propagation and attenuation and their influences on earth's
977 atmosphere. Ph.D. thesis, Yale University, New Haven, Connecticut.

978 Kruse, C. G., 2020: Regional to global evolution of impacts of parameterized mountain-
979 wave drag in the lower stratosphere. *J. Climate*, **33 (8)**, 3093 – 3106, https://doi.org/
980 10.1175/JCLI-D-19-0076.1.

981 Kruse, C. G., and R. B. Smith, 2015: Gravity wave diagnostics and characteristics in mesoscale
982 fields. *J. Atmos. Sci.*, **72 (11)**, 4372–4392, https://doi.org/10.1175/JAS-D-15-0079.1.

983 Kruse, C. G., and R. B. Smith, 2018: Nondissipative and dissipative momentum deposition by
984 mountain wave events in sheared environments. *J. Atmos. Sci.*, **75 (8)**, 2721–2740, https://doi.org/
985 10.1175/JAS-D-17-0350.1.

986 Kruse, C. G., R. B. Smith, and S. D. Eckermann, 2016: The midlatitude lower-stratospheric
987 mountain wave "valve layer". *J. Atmos. Sci.*, **73 (12)**, 5081–5100, https://doi.org/10.1175/
988 JAS-D-16-0173.1.

Lauritzen, P. H., J. T. Bacmeister, P. F. Callaghan, and M. A. Taylor, 2015: Ncar_topo (v1.0): Ncar global model topography generation software for unstructured grids. *Geoscientific Model Development*, **8 (12)**, 3975–3986, https://doi.org/10.5194/gmd-8-3975-2015.

Leuenberger, D., M. Koller, O. Fuhrer, and C. Schär, 2010: A generalization of the sleve vertical coordinate. *Mon. Wea. Rev.*, **138 (9)**, 3683–3689, https://doi.org/10.1175/2010MWR3307.1.

Lindzen, R. S., 1981: Turbulence and stress owing to gravity wave and tidal breakdown. *J. Geophys. Res.: Oceans*, **86 (C10)**, 9707–9714, https://doi.org/10.1029/JC086iC10p09707.

Liu, H., K. C. Jezek, and Z. Zhao, 2015: Radarsat antarctic mapping project digital elevation model, version 2. Tech. rep., NASA National Snow and Ice Data Center Distributed Active Archive Center, https://doi.org/10.5067/8JKNEW6BFRVD, Boulder, Colorado.

Lott, F., and M. J. Miller, 1997: A new subgrid-scale orographic drag parametrization: Its formulation and testing. *Quart. J. Roy. Meteor. Soc.*, **123 (537)**, 101–127, https://doi.org/10.1002/qj.49712353704.

Marsh, D. R., M. J. Mills, D. E. Kinnison, J.-F. Lamarque, N. Calvo, and L. M. Polvani, 2013: Climate change from 1850 to 2005 simulated in cesm1(waccm). *J. Climate*, **26 (19)**, 7372 – 7391, https://doi.org/10.1175/JCLI-D-12-00558.1.

McFarlane, N. A., 1987: The effect of orographically excited gravity wave drag on the general circulation of the lower stratosphere and troposphere. *J. Atmos. Sci.*, **44**, 1775–1800.

McLandress, C., T. G. Shepherd, S. Polavarapu, and S. R. Beagley, 2012: Is missing orographic gravity wave drag near 60^oS the cause of the stratospheric zonal wind biases in chemistry–climate models? *J. Atmos. Sci.*, **69**, 802–818.

Meyer, C. I., M. Ern, L. Hoffmann, Q. T. Trinh, and M. J. Alexander, 2018: Intercomparison of AIRS and HIRDLS stratospheric gravity wave observations. *Atmospheric Measurement Techniques*, **11 (1)**, 215–232, https://doi.org/10.5194/amt-11-215-2018.

Miller, M. J., T. N. Palmer, and R. Swinbank, 1989: Parameterization and influence of subgridscale orography in general circulation model and weather prediction models. *Meteor. Atmos. Phys.*, **40**, 84–209, https://doi.org/10.1007/BF01027469.

Neale, R. B., and Coauthors, 2010: Description of the NCAR Community Atmosphere Model (CAM5.0). Tech. rep., NCAR Tech. Note NCAR/TN-486+STR.

Orr, A., P. Bechtold, J. Scinocca, M. Ern, and M. Janiskova, 2010: Improved middle atmosphere climate and forecasts in the ecmwf model through a nonorographic gravity wave drag parameterization. *J. Climate*, **23 (22)**, 5905 – 5926, https://doi.org/10.1175/2010JCLI3490.1.

Orr, A., and Coauthors, 2015: Inclusion of mountain-wave-induced cooling for the formation of PSCs over the Antarctic Peninsula in a chemistry–climate model. *Atmos. Chem. Phys.*, **15 (2)**, 1071–1086, https://doi.org/10.5194/acp-15-1071-2015.

Palmer, T. N., G. J. Shutts, and R. Swinbank, 1986: Alleviation of a systematic westerly bias in general circulation and numerical weather prediction models through an orographic gravity wave drag parametrization. *Quart. J. Roy. Meteor. Soc.*, **112**, 1001–1039.

Rabier, F., and Coauthors, 2010: The concordiasi project in antarctica. *Bull. Amer. Meteor. Soc.*, **91 (1)**, 69 – 86, https://doi.org/10.1175/2009BAMS2764.1.

Rapp, M., and Coauthors, 2021: Southtrac-gw: An airborne field campaign to explore gravity wave dynamics at the world's strongest hotspot. *Bull. Amer. Meteor. Soc.*, **102 (4)**, E871 – E893, https://doi.org/10.1175/BAMS-D-20-0034.1.

Sato, K., S. Tateno, S. Watanabe, and Y. Kawatani, 2012: Gravity wave characteristics in the southern hemisphere revealed by a high-resolution middle-atmosphere general circulation model. *J. Atmos. Sci.*, **69 (4)**, 1378 – 1396, https://doi.org/10.1175/JAS-D-11-0101.1.

Scinocca, J. F., 2003: An accurate spectral nonorographic gravity wave drag parameterization for general circulation models. *J. Atmos. Sci.*, **60 (4)**, 667 – 682, https://doi.org/10.1175/1520-0469(2003)060<0667:AASNGW>2.0.CO;2.

Scinocca, J. F., and N. A. McFarlane, 2000: The parametrization of drag induced by stratified flow over anisotropic orography. *Quart. J. Roy. Meteor. Soc.*, **126**, 2353–2393, https://doi.org/doi:10.1002/qj.49712656802.

Skamarock, W. C., 2004: Evaluating mesoscale nwp models using kinetic energy spectra. *Monthly Weather Review*, **132 (12)**, 3019 – 3032, https://doi.org/10.1175/MWR2830.1, URL https://journals.ametsoc.org/view/journals/mwre/132/12/mwr2830.1.xml.

Skamarock, W. C., H. Ong, and J. B. Klemp, 2021: A fully compressible nonhydrostatic deep-atmosphere equations solver for mpas. *Mon. Wea. Rev.*, **149 (2)**, 571 – 583, https://doi.org/10.1175/MWR-D-20-0286.1.

Smith, R. B., 2019: 100 Years of Progress on Mountain Meteorology Research. *Meteorological Monographs*, **59**, 20.1–20.73, https://doi.org/10.1175/AMSMONOGRAPHS-D-18-0022.1.

Smith, R. B., and C. G. Kruse, 2017: Broad-spectrum mountain waves. *J. Atmos. Sci.*, **74 (5)**, 1381–1402, https://doi.org/10.1175/JAS-D-16-0297.1.

van Niekerk, A., and S. Vosper, 2021: Towards a more 'scale-aware' orographic gravity wave drag parametrization: description and initial testing. *Quart. J. Roy. Meteor. Soc.*, in press, https://doi.org/https://doi.org/10.1002/qj.4126.

Walters, D., and Coauthors, 2017: The Met Office Unified Model global atmosphere 6.0/6.1 and JULES global land 6.0/6.1 configurations. *Geoscientific Model Development*, **10 (4)**, 1487–1520, https://doi.org/10.5194/gmd-10-1487-2017, URL https://gmd.copernicus.org/articles/10/1487/2017/.

Weimer, M., and Coauthors, 2021: Mountain-wave-induced polar stratospheric clouds and their representation in the global chemistry model icon-art. *Atmos. Chem. Phys.*, **21 (12)**, 9515–9543, https://doi.org/10.5194/acp-21-9515-2021.

Wright, C. J., N. P. Hindley, L. Hoffmann, M. J. Alexander, and N. J. Mitchell, 2017: Exploring gravity wave characteristics in 3-D using a novel S-transform technique: AIRS/Aqua measurements over the Southern Andes and Drake Passage. *Atmos. Chem. and Phys.*, **17 (13)**, 8553–8575, https://doi.org/10.5194/acp-17-8553-2017.

Zängl, G., D. Reinert, P. Rípodas, and M. Baldauf, 2015: The ICON (ICOsahedral Non-hydrostatic) modelling framework of DWD and MPI-M: Description of the non-hydrostatic dynamical core. *Quart. J. Roy. Meteor. Soc.*, **141 (687)**, 563–579, https://doi.org/https://doi.org/10.1002/qj.2378.