SPECIAL

Multi-Scale Dynamics of Gravity Waves (MS-GWaves)

COLLECTION

Estimates of Southern Hemispheric Gravity Wave Momentum Fluxes across Observations, Reanalyses, and Kilometer-Scale Numerical Weather Prediction Model✎

AMAN GUPTA[a], ROBERT REICHERT,[a,b] ANDREAS DÖRNBRACK,[b] HELLA GARNY,[a,b] ROLAND EICHINGER,[b,d] INNA POLICHTCHOUK,[c] BERND KAIFLER,[b] AND THOMAS BIRNER[a,b]

[a] Meteorological Institute, Ludwig Maximilians Universität, Munich, Germany
[b] Institut für Physik der Atmosphäre, Deutsches Zentrum für Luft- und Raumfahrt, Oberpfaffenhofen, Germany
[c] European Centre for Medium-Range Weather Forecasts, Reading, United Kingdom
[d] Department of Atmospheric Physics, Faculty of Mathematics and Physics, Charles University Prague, Prague, Czech Republic

ABSTRACT: Gravity waves (GWs) are among the key drivers of the meridional overturning circulation in the mesosphere and upper stratosphere. Their representation in climate models suffers from insufficient resolution and limited observational constraints on their parameterizations. This obscures assessments of middle atmospheric circulation changes in a changing climate. This study presents a comprehensive analysis of stratospheric GW activity above and downstream of the Andes from 1 to 15 August 2019, with special focus on GW representation ranging from an unprecedented kilometer-scale global forecast model (1.4 km ECMWF IFS), ground-based Rayleigh lidar (CORAL) observations, modern reanalysis (ERA5), to a coarse-resolution climate model (EMAC). Resolved vertical flux of zonal GW momentum (GWMF) is found to be stronger by a factor of at least 2–2.5 in IFS compared to ERA5. Compared to resolved GWMF in IFS, parameterizations in ERA5 and EMAC continue to inaccurately generate excessive GWMF poleward of 60°S, yielding prominent differences between resolved and parameterized GWMFs. A like-to-like validation of GW profiles in IFS and ERA5 reveals similar wave structures. Still, even at ~1 km resolution, the resolved waves in IFS are weaker than those observed by lidar. Further, GWMF estimates across datasets reveal that temperature-based proxies, based on midfrequency approximations for linear GWs, overestimate GWMF due to simplifications and uncertainties in GW wavelength estimation from data. Overall, the analysis provides GWMF benchmarks for parameterization validation and calls for three-dimensional GW parameterizations, better upper-boundary treatment, and vertical resolution increases commensurate with increases in horizontal resolution in models, for a more realistic GW analysis.

SIGNIFICANCE STATEMENT: Gravity wave–induced momentum forcing forms a key component of the middle atmospheric circulation. However, complete knowledge of gravity waves, their atmospheric effects, and their long-term trends are obscured due to limited global observations, and the inability of current climate models to fully resolve them. This study combines a kilometer-scale forecast model, modern reanalysis, and a coarse-resolution climate model to first compare the resolved and parameterized momentum fluxes by gravity waves generated over the Andes, and then evaluate the fluxes using a state-of-the-art ground-based Rayleigh lidar. Our analysis reveals shortcomings in current model parameterizations of gravity waves in the middle atmosphere and highlights the sensitivity of the estimated flux to the formulation used.

Gupta's current affiliation: Doerr School of Sustainability, Stanford University, Stanford, California.

Corresponding author: Aman Gupta, ag4680@stanford.edu

1. Introduction

Gravity waves (GWs) are among the key drivers of the meridional overturning circulation in the mesosphere and upper stratosphere. They significantly contribute to the large-scale variability in the mesosphere by driving the pole-to-pole mass transport (Holton 1982; Fritts and Alexander 2003; Becker 2012). They impact the momentum forcing of key stratospheric processes like the quasi-biennial oscillation (QBO) of tropical stratospheric winds (Giorgetta et al. 2002; Ern et al. 2014), and

the spring breakdown of the Antarctic polar vortex (Gupta et al. 2021). GWs can also contribute to triggering a rapid breakdown of the wintertime polar vortex, i.e., sudden strato-spheric warming (Albers and Birner 2014; Song et al. 2020), and influence tropospheric storm tracks (Kidston et al. 2015; Domeisen and Butler 2020). In addition, GWs couple different layers of the atmosphere by transporting energy and momen-tum from near surface to high altitudes (Fritts and Alexander 2003).

GWs also influence atmospheric transport and chemistry as they propagate and dissipate (Garcia and Solomon 1985; Xu et al. 2000; Eichinger et al. 2020; Weimer et al. 2023, for in-stance). GW-induced cold anomalies in the polar winter strato-sphere provide suitable conditions for the formation of polar stratospheric clouds, enabling reactions that promote the de-struction of ozone (Dörnbrack et al. 1999; Höpfner et al. 2006; Hoffmann et al. 2017). A missing or inaccurate representation of GWs in models contributes to the "cold-pole" bias in climate models, i.e., a later-than-observed westerly-to-easterly transition of the springtime zonal winds in the high-latitude stratosphere (McLandress et al. 2013). The bias is further amplified by the re-sulting feedback between ozone transport and radiation. An ac-curate representation of GWs in climate models is, thus, key to understanding the role played by mesoscales and small scales in driving atmospheric variability and global circulation (Eichinger et al. 2020).

The spatial scales of atmospheric GWs and their sources (e.g., convection, orography, jets, and fronts) can range from $\mathcal{O}(100)$ m to $\mathcal{O}(1000)$ km (Fritts and Alexander 2003). For this reason, GW variability constitutes a "gray zone" for dynamics in global weather prediction and climate models, i.e., the models only resolve part of the GW spectrum at their operational reso-lution (Plougonven et al. 2020). Depending on the grid resolu-tion and complexity, climate models require representation of the underresolved or unresolved part of the GW spectrum through the use of parameterizations, even at a 10 km resolution (Polichtchouk et al. 2022, 2023). The free parameters in the pa-rameterizations often require tuning (Hourdin et al. 2017). However, any robust evaluation is obscured by the limited avail-ability of adequate observations. GW schemes are mostly tuned using wind climatologies, i.e., via "missing drag," rather than via direct comparison to observed GW parameters. An alternate strategy, one explored in this study, would be to tune the param-eterizations against validated high-resolution models. Therefore, advancements in atmospheric GW modeling benefit from both an expanding set of GW observations and improvements in model resolution.

The impact of model resolution on the resolved GW repre-sentation is illustrated in Fig. 1, which shows the vertical mo-mentum flux convergence (VMFC) for August 2019 in three different reanalyses: ERA-Interim, JRA-55, and ERA5. The VMFC represents the zonal GW forcing explicitly resolved by the models and is computed here as the vertical derivative of the vertical flux of zonal momentum, i.e., $-\overline{u'\omega'_p}$, for zonal wavenumbers 21 and higher. The three reanalyses have very different horizontal and/or vertical grid resolutions, but their large-scale winds in the troposphere and the stratosphere are constrained to be nearly identical. The VMFC is strongest for
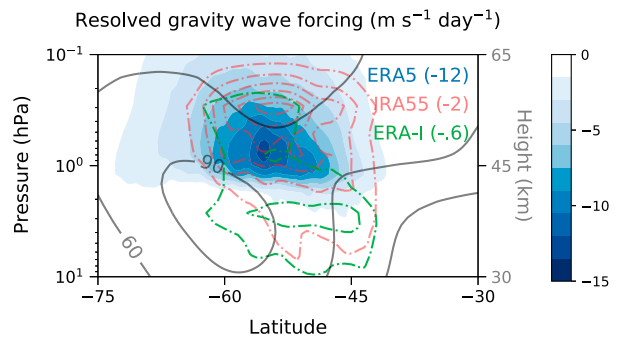

FIG. 1. Mean vertical momentum flux convergence (VMFC) of zonal momentum flux (in m s^{-1} day^{-1}) during 1–15 Aug 2019 across three different reanalyses: ERA5 (color shading), JRA-55 (red), and ERA-Interim (green). The VMFC represents the re-solved GW forcing in the zonal direction and has been computed as the covariance $-\overline{u'\omega'_p}$ for zonal wavenumbers 21 and higher. The black contours show the zonal-mean zonal wind (in m s^{-1}) during this period in ERA5. Contour intervals for both JRA-55 and ERA-Interim are 0.3 m s^{-1} day^{-1}. The twin vertical axis shows the mean geometric height for reference. The numbers in the pa-rentheses show the maximum forcing value (in m s^{-1} day^{-1}) over all pressure levels and latitudes shown.

ERA5 (maximum: -12 m s^{-1} day^{-1}), and roughly an order of magnitude lower for the coarser-resolution JRA-55 (maximum: -2 m s^{-1} day^{-1}) and ERA-Interim (maximum: -0.6 m s^{-1} day^{-1}). As will be shown later, even ERA5 with a horizontal resolution of 30 km only resolves less than 40% of the actual forcing.

In lieu of GW-resolving models and reanalyses, much of our understanding of the GW source spectrum and flux distribution comes through in situ measurements and remote soundings of the atmospheric radiance temperature, horizontal velocity, and trace gas profiles. Remote soundings of GWs us-ing satellite-based infrared limb sounders, nadir instruments, and radio occultation devices, have empowered the identifica-tion of GW hotspots globally (Hertzog et al. 2008; Schroeder et al. 2009; Alexander and Grimsdell 2013; Hoffmann et al. 2013; Ern et al. 2018; Hindley et al. 2020; Banyard et al. 2021). Interannual observational records from ground-based lidars and radars over multiple fixed locations have further advanced our knowledge of the seasonal cycle of GW fluxes over these hot-spots (Sato et al. 2009; Dörnbrack et al. 2017; Kaifler et al. 2020; Minamihara et al. 2020; Reichert et al. 2021). Atmospheric soundings of GWs using airborne lidar (Rapp et al. 2021, for in-stance), along with trajectories provided by superpressure bal-loons (Hertzog et al. 2008; Rabier et al. 2010; Corcos et al. 2021; Lindgren et al. 2020) have improved our understanding of the GW spectrum and intermittency, and their spatial variations, bridging theoretical understanding with observational validation.

These new, multifidelity observations provide ever-growing measurements of atmospheric GWs but they are still either limited in their spatial coverage or observe GWs over limited frequency bands (Geller et al. 2013). For instance, super-pressure balloons can measure GWs in the horizontal, but strug-gle to capture their vertical profiles. In contrast, ground-based lidars capture GW vertical profiles well but do not capture their

horizontal profiles. Further, spatial scales crucial for GW evolution are often much finer than the observational filter of satellite instruments, like HIRDLS or AIRS. Even state-of-the-art satellite instruments cannot measure vertical velocity, which is necessary to directly compute the GW pseudomomentum flux. Such limiting factors make comparison of modeled and observed GWs challenging and obscure a comprehensive understanding of GW dynamical interaction in the atmosphere [see Alexander et al. (2010) for details].

Model parameterizations of GWs, too, present themselves as one of the leading sources of uncertainty in stratosphere-resolving climate models (van Niekerk et al. 2020; Plougonven et al. 2020; Mansfield and Sheshadri 2022). Most operational orographic and nonorographic GW parameterizations are subject to simplifying assumptions of strictly vertical propagation, infinite propagation speed, and an a priori source spectrum for launching nonorographic GWs; all of these assumptions partially violate the insights gained from theory and observations over the past several decades (Kim et al. 2003).

Limited observations, parameterization uncertainty, and increasing computational power motivate the use of high-resolution numerical weather prediction (NWP) models—which explicitly resolve a major portion of the GW spectrum—to understand atmospheric GWs. NWP models enable a detailed simulation of atmospheric GWs and permit model validations using high-resolution observations (Alexander and Teitelbaum 2011; Shutts and Vosper 2011; Sato et al. 2012; Preusse et al. 2014; Hindley et al. 2021; Kruse et al. 2022; Gisinger et al. 2022; Polichtchouk et al. 2022, 2023; Procházková et al. 2023). As a recent example, the unprecedented 1.4 km global mesoscale-resolving runs using European Centre for Medium-Range Weather Forecasts (ECMWF)'s Integrated Forecast System (IFS) (Wedi et al. 2020; Polichtchouk et al. 2022, 2023) have provided new insights into GW flux contribution to tropical and extratropical dynamics from different scales: wavelengths shorter than 10 km, between 10 and 100 km, and longer than 100 km.

This study predominantly focuses on GWs over the Andes during the 1–15 August 2019 period. The Andes are one of the strongest GW hotspots on the planet (Rapp et al. 2021). Under the effects of rotation, GWs generated here can be found downstream of the Andes under long-lasting cross-mountain flows. These waves contribute to the momentum budget around 60°S over the Southern Ocean (Queney 1948; Gill 1982; Sato et al. 2012). In fact, GW activity from these waves can be detected even halfway around the 60°S latitude circle, alongside GW activity from other orographic features and storm tracks (Hendricks et al. 2014; Gupta et al. 2021). Missing or inaccurate model representation of GW fluxes in this region presents as a key source of circulation bias in stratosphere-resolving climate models (McLandress et al. 2013).

We use an array of datasets to analyze GWs representation and momentum forcing both around the Andes, and globally. We blend the 1.4 km IFS data, with observations, reanalysis, and a coarse climate model to obtain estimates of gravity wave momentum flux (GWMF) in the extratropical austral stratosphere. The datasets complement each other by differing in their GW representation: ground-based lidar observations measure the true state of the atmosphere, the kilometer-scale IFS pretty much resolves the complete mesoscale GW spectrum, the reanalysis partly resolves and partly parameterizes mesoscale GWs, and the nudged coarse-grid climate model exclusively parameterizes all GW effects. To our knowledge, such an analysis covering the whole range of resolved plus parameterized GW representations has not been performed within a single study before, and has multiple merits:

 (i) It facilitates the first validation of a kilometer-scale global model using remote sensing observations.
(ii) "Model truth" from the kilometer-scale IFS provides a baseline benchmark for GW fluxes against which to evaluate GW representation skill of modern reanalysis. It also promotes an evaluation of the strengths and deficiencies of parameterized GW representations in coarse climate models. Consequently, it allows analyzing the impact of imperfect GW parameterization on the modeled stratospheric circulation.
(iii) Augmenting lidar observations with the validated IFS and ERA5 allows blending the local observational estimates of GWMF vertically above Rio Grande to the regional and global GW flux in both GW resolving and GW parameterizing models. This provides usable benchmarks to inform model parameterizations.
(iv) Models simulate global temperature and winds, but most satellites and lidars still only measure temperature. Combining validated IFS with lidar and reanalysis allows connecting two very different ways of estimating GWMFs: one using winds and one using temperature.

Therefore, with this analysis, we essentially lay out a strategy for a comprehensive model evaluation: using observations to validate high-resolution global models, and subsequently using them as "model truths" to set benchmarks for testing lower-resolution climate models.

Our analysis complements the regional treatise by Kruse et al. (2022), which employs a suite of 3 km regional and 10 km global NWP models to study mountain waves locally around the Andes and over the Drake Passage. This study, however, focuses on a different time period during peak winter, using a different set of models, reanalyses, and observations, and has a broader focus on estimating GW fluxes and forcing over multiple spatial scales: pointwise, regional, and global. Our analysis follows the natural next step by employing a ~1 km global model to study GWs in the Southern Hemisphere. In addition, our analysis also investigates the effectiveness and challenges when using linear GW theory to connect GWMF estimates from models versus ground-based observations.

The details of the different datasets employed in the analysis are provided in section 2. The methodology to compute GWMFs and other diagnostics used in the analysis is discussed in section 3. The results are presented in section 4, where we compare the resolved and parameterized GW forcing, assess the horizontal and vertical GW profiles in IFS and ERA5, and wherever possible, validate them with observations from the ground-based lidar [Compact Rayleigh Autonomous Lidar (CORAL)]. Ultimately, we provide GWMF estimates from observational and model data

products using two different approximations of the pseudomomentum flux. Finally, we discuss our findings and summarize our conclusions in section 5.

2. Models and datasets

a. The Compact Rayleigh Autonomous Lidar

CORAL is a Rayleigh backscatter lidar designed for measurements of middle atmospheric temperature up to an altitude of 100 km (Kaifler and Kaifler 2021). It was built by the German Aerospace Center (DLR) and installed at the Estación Astronómica Río Grande (EARG) in Rio Grande (53.79°S, 67.75°W), Argentina, in late November 2017 (Fig. 2, violet marker). CORAL has since been recording temperatures in the middle atmosphere over Rio Grande (Reichert et al. 2021). Large-amplitude waves observed over Rio Grande can have long wavelengths associated with the hydrostatic, rotating wave regime, with a nonzero leeward component of group velocity (Queney 1948). So even while propagating perfectly against the mean flow, these waves carry energy downstream of the Andes (Reichert et al. 2021). This lateral and leeward propagation of GWs permits recording these waves over Rio Grande, which by itself is a relatively flat region southeast of the Andes.

CORAL probes the atmosphere whenever the night sky is clear and allows for investigation of variations in temperature and GW activity in the middle atmosphere on hourly to seasonal time scales. For the period considered in this study, i.e., the first half of August 2019, CORAL operated over 6 nights: 2, 4, 5, 7, 8, and 9 August 2019, and collected over 49 h of temperature data over Rio Grande (as shown in Fig. 7a). Temperature data in the altitude range from 15 to 90 km are provided on a 5 min × 100 m grid (Kaifler and Kaifler 2021).

To simplify comparison with other datasets, the CORAL temperature profiles are reduced in number to a subset of profiles that are at most 3 min off the full hour. In addition, a running mean over 2 km is applied to the vertical temperature profiles to mimic the coarser ERA5 resolution.

b. 1.4 km ECMWF IFS runs

The ECMWF IFS run considered in this study is identical to one of the two climate model simulations described and used in Polichtchouk et al. (2022, 2023). The simulation was initialized at 0000 UTC 1 August 2019 from the ECMWF operational analysis and integrated with a time step size of 60 s at approximately a 1.4 km global grid spacing using 7999 spherical harmonics. The unprecedented TCo7999 horizontal-resolution simulation was performed on the Summit supercomputer, accessed through an Innovative and Novel Computational Impact on Theory and Experiment award (INCITE; Wedi et al. 2020). While the free-running simulation was integrated for four months, here we focus on the first 15 days of the simulation when the background flow and temperature distribution in the simulation are similar to (re)analyses. Especially for the first 5 days of the simulation, Polichtchouk et al. (2022, 2023) verified that the background flow and temperature structure are practically indistinguishable from reanalyses. Hereafter, we interchangeably refer to this simulation as "IFS-1km."
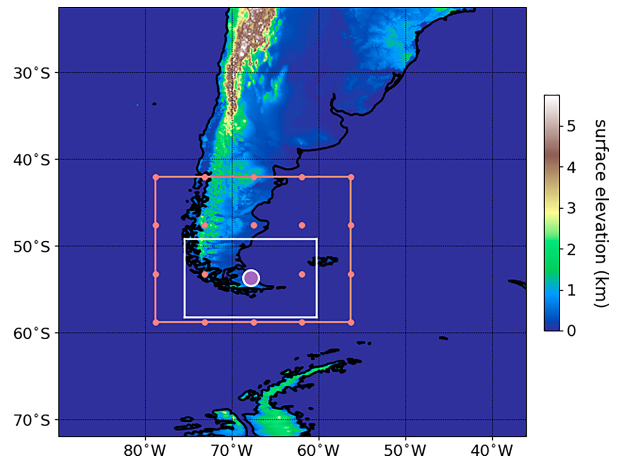

FIG. 2. Variations in surface topography are shown in color using cylindrical projection. The violet marker shows the location of CORAL lidar at Rio Grande. The white box shows the 1000 km × 1000 km box around Rio Grande used for two-dimensional wavelet analysis. The red grid box and dots shows the T21 grid used to conservatively interpolate high-resolution fluxes from IFS-1km and ERA5.

The simulation was performed with the full-complexity global semi-implicit semi-Lagrangian spectral ECMWF IFS atmosphere model (based on cycle 45r1; IFS 2018) and forced by the 0.05° OSTIA sea surface temperature and sea ice data (Donlon et al. 2012). The IFS is discretized in the horizontal using a spherical harmonic expansion and a cubic-octahedral grid.

In the vertical, IFS is discretized using a third-order vertical finite element scheme on a pressure-based hybrid coordinate using 137 vertical levels extending from the surface up to 0.01 hPa. The vertical level spacing (in m) increases toward the model top: for instance, at 1000 hPa it is ∼10 m; at 500 hPa ∼100 m, at 100 hPa ∼300 m, at 30 hPa ∼500 m, and at 10 hPa ∼700 m.

To prevent spurious wave reflection at the model top, a fourth-order hyperdiffusion (∇^4) is applied on vorticity, divergence, and temperature fields above 10 hPa to damp vertically propagating waves with a scale-selective and height-dependent damping time scale [units: (s)] roughly proportional to the fourth power of the ratio of the maximum wavenumber, $N_{max} = 7999$ and the wavenumber, n, i.e., $\propto (N_{max}/n)^4$. This hyperdiffusion is quite weak and has a small impact on the resolved waves. In addition, a first-order diffusion (∇) is applied on the divergence field above 1 hPa with a damping time scale roughly proportional to the ratio N_{max}/n. This diffusion is very strong and very effective at damping all resolved waves. Any analysis of resolved GWs in the mesosphere will, therefore, be dominated by the sponge effects and thus we restrict most of our GW analysis up to the stratopause only.

Both the deep convection and gravity wave parameterizations are switched off for the 1.4 km simulation; all contributions to the VMFC, i.e., the wave drag, come exclusively from resolved waves. The model resolves the whole mesoscale GW spectrum but resolves only a fraction of the submesoscale

spectrum. Gridscale hyperdiffusion and other numerical method choices in the model reduce its effective grid resolution from $\Delta x = 1.4$ km to about $6\Delta x$–$8\Delta x$ (Skamarock 2004; Klaver et al. 2020). Due to data management issues with the model winds during postprocessing, we consider the GW effects on the zonal momentum budget only. Hence, we use 3-hourly instantaneous fields on model levels to calculate the zonal momentum forcing due to resolved GWs as described in section 3.

For a more detailed discussion on the model and GW momentum flux computation in IFS, see Polichtchouk et al. (2022, 2023).

c. ERA5

We use the fifth generation reanalysis from ECMWF, ERA5 (Hersbach et al. 2020), in our analysis. ERA5 winds and temperature are available for 1940 to the present on a $0.3° \times 0.3°$ uniform latitude–longitude grid (~30 km) in the horizontal and 137 model levels in the vertical. The vertical configuration is identical for both ERA5 and IFS-1km.

The vertical levels and the analytical definition for model top damping in ERA5 are identical to those in IFS. However, ERA5 uses ECMWFs IFS 2016, cycle 41r2, to generate the reanalysis output, and uses a total of 639 spherical harmonics, i.e., $N_{max} = 639$. Therefore, waves of any given scale are damped more strongly in ERA5 (shorter damping time) than they are in the IFS-1km model, leading to differences in small-scale damping above 1 hPa.

At a horizontal grid resolution of 30 km, with added effects from gridscale hyperdiffusion, ERA5 resolves GWs with horizontal wavelength ~200 km and longer (Skamarock 2004; Gupta et al. 2021; Pahlavan et al. 2023) and parameterizes the rest. The unresolved part of the GW spectrum is parameterized using the Lott and Miller (1997) scheme for orographic GWs and the Scinocca (2003) scheme for nonorographic GWs. The exact configuration of the nonorographic GW scheme in ERA5 is discussed in Orr et al. (2010). In general, the difference between analysis winds and forecast winds in ERA5, i.e., the assimilation increment, can also inform of missing GW forcing by informing about the deviation of the free-running model from the true assimilated atmospheric state (Sato and Hirano 2019). We find the contribution from the analysis increment to be small for the 1–15 August 2019 period (online supplementary Fig. S1) and, therefore, do not include it in our analysis.

We use hourly temperature and winds for 1–15 August 2019 to compute the GW momentum fluxes over the period. In addition, we use 6-hourly temperature and winds for June, July, and August over the 20-yr period of 2000–19 for a climatological momentum flux analysis.

d. EMAC chemistry–climate model

ECHAM MESSy Atmospheric Chemistry (EMAC, v2.55.0; Jöckel et al. 2010, 2016) model data were taken from the so-called ref-D1SD simulation, which was conducted for the IGAC/SPARC Chemistry-Climate Model Initiative 2022 (CCMI-2022).

CCMI-2022 is the successor of the Chemistry-Climate Model Initiative phase 1 (CCMI-1), the simulation setups of which are documented in detail in Morgenstern et al. (2017). The ref-D1SD simulation denotes a hindcast simulation spanning the period 1979–2020. This is a simulation with specified dynamics (SD), using ERA5 reanalysis data (see above). This means that surface pressure, vorticity, divergence, and temperature are nudged toward ERA5 by Newtonian relaxation every hour. The last three quantities are nudged from the surface up to ~100 hPa. For the simulation, emissions are taken from MACCity (Granier et al. 2011) and prescribed sea surface temperatures and sea ice from HadISST are used (Rayner et al. 2003).

The EMAC simulation was integrated at a T42 horizontal resolution of $\sim2.8° \times 2.8°$ in latitude and longitude of the corresponding quadratic Gaussian grid, with 90 hybrid layers in the vertical and explicitly resolved middle atmosphere dynamics (T42L90MA). In this setup, the vertical resolution in the upper troposphere–lower stratosphere region (UTLS) is ~500–600 m and the uppermost model layer is centered at ~0.01 hPa.

The nonorographic GW module GWAVE (Baumgaertner et al. 2013) uses the parameterization originally developed by Hines (1997). The launch level where GWs are released is set to be near 643 hPa and the namelist parameter *rmscon* is set to 0.92 to achieve an optimal strength of the Antarctic polar vortex (see Jöckel et al. 2016). The rmscon parameter refers to the root-mean-square of the nonorographic GW-induced wind speed (in m s^{-1}) at the launch level and controls the momentum deposition in the stratosphere and mesosphere, thus modulating features like the QBO and vortex strength. Orographic GWs in EMAC are parameterized in the submodel OROGW using the columnar approach by Lott and Miller (1997), and its use in ECHAM is described in Roeckner et al. (2003).

3. Theory and methodology

a. Identifying GWs in the upper stratosphere

To identify days with notable GW activity over Rio Grande, Argentina (53.69°S, 67.75°W), we define a GW temperature variance index β (units: K^2) as

$$\beta(t) = \frac{1}{\Delta p} \int_{p_1}^{p_2} T'^2 dp, \qquad (1)$$

where p is pressure, $p_1 = 1.5$ hPa, $p_2 = 10$ hPa, $\Delta p = p_2 - p_1$, and T' is the small-scale temperature perturbation above Rio Grande obtained from 6-hourly ERA5 data by high-pass filtering of zonal wavenumbers 21 and above. Simply put, β captures the average small-scale temperature variance in the upper stratosphere over a fixed point over Rio Grande. In this study, we treat $10 \leq \beta \leq 100$ K^2 (which roughly corresponds to a mean wave amplitude between ~3 and 10 K) as a moderately strong GW, and $\beta \gtrsim 100$ K^2 as a strong GW (with a mean amplitude of 10 K and above).

b. Two formulations for the GWMF

GWs propagating conservatively (without dissipation) through a vertically varying flow conserve wave pseudomomentum (Fritts

and Alexander 2003). The associated vertical flux of horizontal GW pseudomomentum may be expressed in terms of Reynolds stresses in pressure coordinates as

$$\mathbf{F} = g^{-1}\left(1 - \frac{f^2}{\hat{\omega}^2}\right)(\overline{u'\omega'}, \overline{v'\omega'}), \tag{2}$$

where u and v are the zonal and meridional winds, ω is the vertical velocity (in Pa s^{-1}) on pressure levels, f is the Coriolis parameter, $\hat{\omega}$ is the intrinsic frequency, the overbar denotes averaging over single or multiple wave cycles (even a zonal mean), and primes denote deviation from the background flow (Fritts and Alexander 2003).

For rotational hydrostatic GWs with frequency moderately larger than the inertial frequency f, i.e., $|f| \ll \hat{\omega}$, the frequency factor, $(1 - f^2/\hat{\omega}^2) \approx 1$, may be neglected (in our case, $\hat{\omega}$ is mostly at least 4–5$|f|$, and hence $f^2/\hat{\omega}^2 < 0.06$), and the GW pseudomomentum may be approximated as the vertical flux of horizontal momentum by GWs:

$$\mathbf{F} = g^{-1}(\overline{u'\omega'}, \overline{v'\omega'}). \tag{3}$$

The true vertical flux of horizontal GW pseudomomentum in Eq. (2) can be computed from model data by considering the contributions from both the vertical flux of zonal momentum and the meridional heat flux due to GWs, i.e., the two components of the vertical Eliassen–Palm flux (Eliassen and Palm 1961). Therefore, neglecting the frequency factor in Eq. (2) is equivalent to neglecting the contribution of heat fluxes due to GWs. The vertical derivative of the zonal component $(\overline{u'\omega'})$ (without the factor g) with respect to pressure p, i.e., $-\partial_p(\overline{u'\omega'})$, will be referred to as the VMFC in the following. It approximates the net zonal-mean zonal forcing due to resolved GWs.

According to Ern et al. (2004), in the midfrequency approximation, i.e., $|f| \ll \hat{\omega} \ll |N|$, $|N|$ being the buoyancy frequency, the dispersion relation for hydrostatic GWs can be simplified to $\hat{\omega}^2 = N^2 k_h^2/m^2$, and can be used to *approximate* the pseudomomentum flux in Eq. (2) using temperature perturbations as

$$\mathbf{F} \approx -2\overline{\rho}E_p\frac{\mathbf{k}_h}{m}, \tag{4}$$

where $\overline{\rho}$ is the background density, $\mathbf{k}_h = (k, l)$ and m are horizontal and vertical wavenumbers, respectively (Ern et al. 2004). Here, $k = 2\pi/\lambda_x$ and $m = 2\pi/\lambda_z$, which implies $k/m = \lambda_z/\lambda_x$. Also, E_p is the GW potential energy per mass and is defined as

$$E_p = \frac{1}{2}\frac{g^2}{N^2}\overline{\left(\frac{T'}{T_0}\right)^2}, \tag{5}$$

where g is acceleration due to Earth's gravity, N is the Brunt–Väisälä frequency, T' is the wave-induced temperature perturbation, and T_0 is the background temperature. The overbar in Eq. (5) denotes an average over single or multiple wave cycles. So far, this has been the most frequently employed formulation to estimate the pseudomomentum flux when only temperature observations are available (Ern et al. 2004, 2014, 2017, 2018; Kaifler et al. 2020; Reichert et al. 2021; Hindley et al. 2020).

We approximate the simplified wave pseudomomentum flux estimated from winds and temperature by respectively computing the vertical flux of zonal momentum in Eq. (3) and the temperature-based proxy in Eq. (4). Neglecting the frequency factor while computing wind covariance neglects a minor (<3%–4%) fraction of the vertical momentum flux accounted for by the heat flux due to GWs (Fig. S2), and using the midfrequency approximation while computing the temperature-based proxy neglects the appropriate scaling for low-frequency GWs, leading to overestimated fluxes from both estimates. Ern et al. (2004) found the midfrequency approximation to lead to a deviation (overestimation) of around ~10% for CRISTA observations.

In this study, we only focus on the zonal component of the GWMF codified in Eqs. (3) and (4), which contributes to the forcing of the zonal flow. Hereafter, we refer to the actual flux computed using Eq. (3) as "direct flux" or simply "flux" (since it represents the vertical momentum flux due to waves), and the flux estimated using temperature variance in Eq. (4) as the "temperature-based flux" or "temperature-based proxy." We remark that, essentially, both the "flux" and the "temperature-based proxy" are two ways to approximate the vertical flux of horizontal wave pseudomomentum.

Recently published Kruse et al. (2022) and Procházková et al. (2023) demonstrate the importance of considering other covariances to completely quantify GW dissipation. These terms include, for instance, the meridional convergence of horizontal momentum, $\overline{u'v'_\phi}$, and zonal convergence of zonal momentum, $\overline{u'u'_\lambda}$, where ϕ and λ are the latitude and longitude, respectively, and subscripts denote partial derivatives [see Procházková et al. (2023) for details]. Depending on the wave event, these additional terms may or may not contribute to the total GW forcing, but should be considered for completeness. For the period analyzed in this study, we find the contribution from these covariances to be an order weaker to VMFC (Figs. S3 and S4). Therefore, hereon, we only focus on VMFC.

c. Computing the direct flux and forcing (VMFC)

To compute the GW forcing in IFS-1km and ERA5 for 1–15 August 2019, we use the procedure described in Polichtchouk et al. (2023). In fact, the IFS zonal GWMFs used in this study are identical to theirs. The full u, v, and ω, were truncated at a T21 truncation by removing the first 20 total wavenumbers. That is, in our computations $u' = u - u_{T21}$ and so on. The quantities were then substituted in Eq. (3) to obtain the wind-based momentum flux, $u'\omega'$. Computing the fluxes in IFS required first removing the truncated large-scale winds from the full velocity field, and then interpolating the product from an O8000 cubic octahedral reduced Gaussian grid to a $0.2° \times 0.2°$ latitude–longitude grid. The product was finally interpolated onto a T21 Gaussian grid by conservatively averaging over the largest retained wavelength, i.e., total wavenumbers 21 and above. For more details on these considerations, see Polichtchouk et al. (2022, 2023).

The GWMF in Polichtchouk et al. (2022, 2023) were interpolated from an O8000 cubic octahedral grid to a T21 Gaussian grid using ECMWF's Meteorological Interpolation and

Regridding (MIR) package (Maciel et al. 2017). For ERA5, we first computed the fluxes on the $0.3° \times 0.3°$ latitude–longitude grid by filtering out the first 21 total wavenumbers, and then coarse grained them to a T21 Gaussian grid using CDO's first-order conservative bilinear interpolation function, *remapcon* (Schulzweida 2005). This way, we conservatively computed the gridbox-averaged flux by averaging over the largest-resolved wavelengths. The T21 grid points locally around Rio Grande are shown in red in Fig. 2.

Our choice of a fixed horizontal wavenumber cutoff to obtain GW temperature perturbations is similar to the Fourier decomposition-based filtering used in Kruse and Smith (2015) and Gisinger et al. (2017). It should be noted that, in principle, since GW characteristics are latitude and height dependent, there is no universal constant cutoff that works globally (Procházková et al. 2023). Recent studies have alternatively used variable cutoffs inferred from the slopes of kinetic energy spectrum (Morfa and Stephan 2023; Stephan et al. 2022) or modal decomposition (Žagar et al. 2018) to differentiate the small scales from large scales. In fact, the estimated cutoffs tend to vary even between different high-resolution models (Stephan et al. 2022), and all these methods still bear sensitivity to height (Procházková et al. 2023). The vertical flux terms, including those considered in this study, however, are less prone to such sensitivities as do the horizontal flux terms whose contribution was found to be an order weaker for the period considered here.

d. Computing the temperature-based proxy

Computing the temperature-based proxy using Eq. (4) requires more methodical steps than computing the flux directly, as it also requires computing both the background stratification and the characteristic horizontal and vertical wavelengths for the GW envelope. In this subsection, we describe the procedure followed to obtain each of these quantities from models and observations and how they were put together to compute the temperature-based proxy.

1) OBTAINING TEMPERATURE PERTURBATIONS

Multiple studies (for instance, Kaifler et al. 2020; Gupta et al. 2021; Polichtchouk et al. 2022) have used a fixed wavenumber cutoff-based filtering to retrieve GWs from the model output temperature. However, vertical profilers like CORAL do not provide any information in the horizontal and must be filtered either vertically or in time (Wilson et al. 1991; Kaifler et al. 2015; Ehard et al. 2015). Hence, to vertically filter the temperature, we apply a high-pass fifth-order Butterworth filter described in Ehard et al. (2015) to retrieve point-based temperature perturbations over Rio Grande in CORAL, IFS, and ERA5. Butterworth filtering effectively separates GWs from the planetary-scale variability dominating the southern midlatitudes, by leveraging the scale separation between the planetary scales and the mesoscales (Fig. S5). As has been highlighted by past studies (Rapp et al. 2018; Strube et al. 2020, for instance), such effectiveness might not be generally applicable elsewhere. To remove stationary structures around the tropopause and stratopause, the 15-day-averaged time-mean

temperature profile (T21-truncated temperature profile for models) was subtracted from the full temperature fields before Butterworth filtering. Doing this effectively filters out the stationary structures without notably impacting the retrieved GW perturbations in the stratosphere (Fig. S6).

To ensure maximum consistency between vertical Butterworth filtering versus fixed wavenumber (T21) horizontal filtering in models, we tested Butterworth filtering with different vertical wavelength cutoffs in IFS-1km. We found minimum RMS error between the horizontal and vertical filtering, for 1–15 August 2019, for a vertical cutoff wavelength of 20 km (Fig. S7). We chose this vertical cutoff to filter out the large-scale temperature in the three datasets and obtain the small-scale temperature perturbation [the term T' in Eq. (4)]. To obtain T' directly over Rio Grande in ERA5, the temperature from the nearest grid points on the $0.3° \times 0.3°$ grid was linearly interpolated to Rio Grande's latitude and longitude.

The filtered-out large-scale temperature was used as the background-temperature profile [the term T_0 in Eq. (4)]. This background temperature T_0 was also used to compute the buoyancy frequency N. We combined these terms to compute the GW potential energy E_p, closely following the approach proposed by Reichert et al. (2021).

2) OBTAINING HORIZONTAL AND VERTICAL WAVELENGTHS

The horizontal wavelengths in IFS and ERA5, and vertical wavelengths in CORAL, IFS, and ERA5 were computed by respectively using a two-dimensional continuous wavelet transform (CWT) around Rio Grande and a one-dimensional CWT in the vertical (Torrence and Compo 1998). To reduce edge effects, the 2D CWT is performed over a 1000 km × 1000 km domain centered around Rio Grande. The domain and the location of Rio Grande are shown using a white box and a location marker, respectively, in Fig. 2. Sample results from the spectral analysis for 4 August 2019 are illustrated in Fig. 3.

The horizontal spectrum for a sample on 4 August 2019 in ERA5 (Fig. 3a) shows that the power contained in the gravity wave packet over Rio Grande is concentrated within two scales: first, around 600 km, where a weak southwest orientation of 5°–15° suggests the wave predominantly propagates in the zonal direction, and the other, around 1500 km, with a southwest tilt of 45°. Here, the orientation is computed as the angle between the total wavenumber and the zonal wavenumber, with an orientation of 0° meaning west and an orientation of 90° meaning south. At each altitude, the power spectrum is normalized by its sum. Subsequently, the normalized spectrum is used to compute a weighted average zonal wavenumber which is considered the zonal wavelength of the wave packet (λ_x). Similarly, the 1D wavelet power spectrum computed from temperature perturbations observed by CORAL on 4 August (Fig. 3b) shows the power to be centered around $\lambda_z = 14$ km, with a half-maximum width of 5 km. The normalized 1D spectrum is weighted averaged to compute the effective vertical wavelength of the wave packet (λ_z). The vertical resolution of all the datasets (~900 m for lidar and ~1.5 km for ERA5 and IFS) is much finer than the estimated vertical
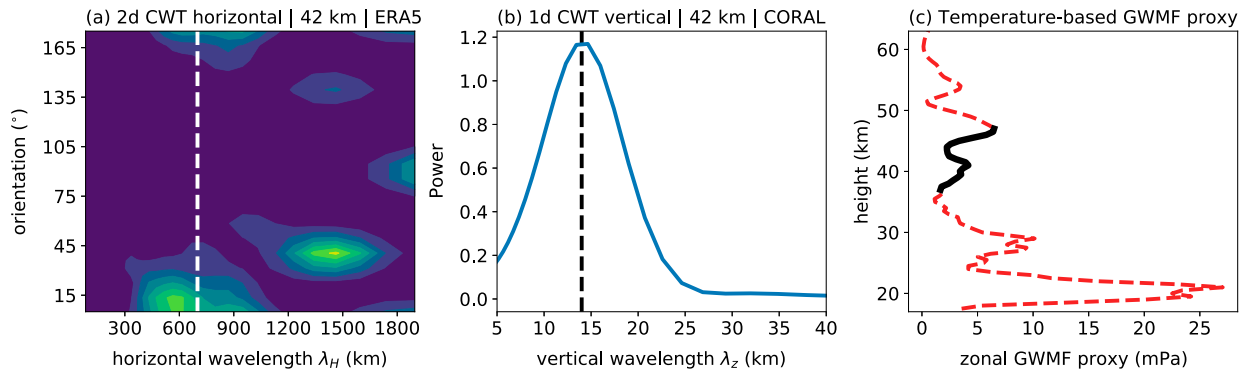
FIG. 3. Figure illustrating the GW spectral properties required to compute the temperature-based proxy. (a) The 2D wavelet power spectrum of the temperature perturbations in ERA5 for 4 Aug 2019 at 42 km altitude. The dashed white vertical bar shows the mean zonal wavelength $\lambda_x = 700$ km. (b) The power spectrum of the temperature perturbations for 4 Aug 2019 at 42 km altitude. The dashed black vertical bar shows the mean vertical wavelength $\lambda_z = 14$ km. (c) The vertical profile of the temperature-based zonal momentum flux proxy for all levels in the stratosphere for 4 Aug 2019. The solid black curve in (c) marks the 37 to 47 km height interval over which the flux was integrated to obtain the fluxes compared in Fig. 11.

wavelengths for all times. Thus, even though limited vertical resolution in models might limit contributions from shorter vertical wavelengths, they well resolve the waves close to the dominant wavelength. If the phase shift was <90°, the models were assumed to approximately reproduce a GW observed by CORAL. Small phase differences are found on 5, 7, and 9 August, while the phase differences are larger on 4 and 8 August.

Since CORAL measures the temperature only over Rio Grande, we use ERA5 to infer the horizontal wavelength of the wave and use that as a substitute for CORAL. This assumption is based on very similar vertical wave profiles among CORAL and ERA5 (shown later in Fig. 7a). A similar approach was adopted by Kaifler et al. (2020) to estimate fluxes from CORAL over Rio Grande for June 2018.

3) COMPUTING THE FLUX FROM THE PARAMETERS

The procedure described above was repeated for all vertical levels in the stratosphere, and along with the GW temperature perturbations (within E_p) was used to compute the temperature-based proxy for all heights (Fig. 3c). The procedure was repeated for all the available time records in CORAL, ERA5, and IFS-1km, from 1 to 15 August 2019. To obtain the proxy at 42 km, the flux was averaged over a 10 km height interval from 37 to 47 km. This interval choice ensured no averaging over the mesospheric sponge (which starts at ~48 km) and averages over a wavelength window that works for most days (10–15 km). We tried multiple such windows around 42 km and found little difference in the fluxes.

4. Results

a. Comparing gravity wave fluxes across IFS-1km, ERA5, and EMAC

We first compare GW fluxes among IFS-1km, ERA5, and EMAC. The three models differ in their GW representation, in that, the mesoscale GWs are completely resolved in IFS-1km,

partly resolved plus partly parameterized in ERA5, and completely parameterized in EMAC.

We focus on the 3–10 August 2019 interval—a period of strong GW excitation over the Andes in the Southern Hemisphere during which IFS-1km and ERA5 barely diverge—and compare the zonal GWMF among the models and reanalysis in the stratosphere at two different altitudes: 25 km (30 hPa; middle stratosphere) and 40 km (3 hPa; upper stratosphere) (Fig. 4). Focusing first on the resolved GWMF in IFS-1km and ERA5 (solid curves), the resolved fluxes demonstrate very similar features and peak between 50° and 55°S at both 25 and 40 km. However, at both altitudes, the resolved fluxes in ERA5 are less than half as strong as the fluxes in IFS-1km, highlighting the impact of resolution. Upon including the parameterized fluxes from ERA5 and EMAC (dashed curves), we note a considerably stronger contribution from parameterizations poleward of 60°S, to the extent that between 60° and 75°S the total fluxes in EMAC and ERA5 become a factor of 2–3 higher, than the fluxes in IFS-1km. This is consistent with Polichtchouk et al. (2023), who found the parameterized nonorographic GWMF in the Southern Hemisphere polar region in IFS at 9 km resolution to be much stronger than the resolved GWMF in IFS at 1 km resolution. Intercomparison efforts by Geller et al. (2013), too, found notably stronger parameterized GWMFs in polar regions for both hemispheres during both summer and winter.

While it is not straightforward to decompose the resolved fluxes in IFS-1km and ERA5 into orographic and nonorographic components, separately analyzing the orographic and nonorographic parameterized fluxes from EMAC suggests that, for the GW case analyzed in this study, more than three-fourths of the parameterized flux in the midlatitudes at 25 km altitude, for latitudes with orography, is attributed to orographic fluxes (Fig. 4b). This fraction is much lower at 40 km (Fig. 4a) due to progressive flux dissipation between 25 and 40 km. Of course, this fraction can vary with the latitude and period under consideration [Fig. 7 of Hertzog et al. (2008) and Fig. 12 of Plougonven et al. (2013)].
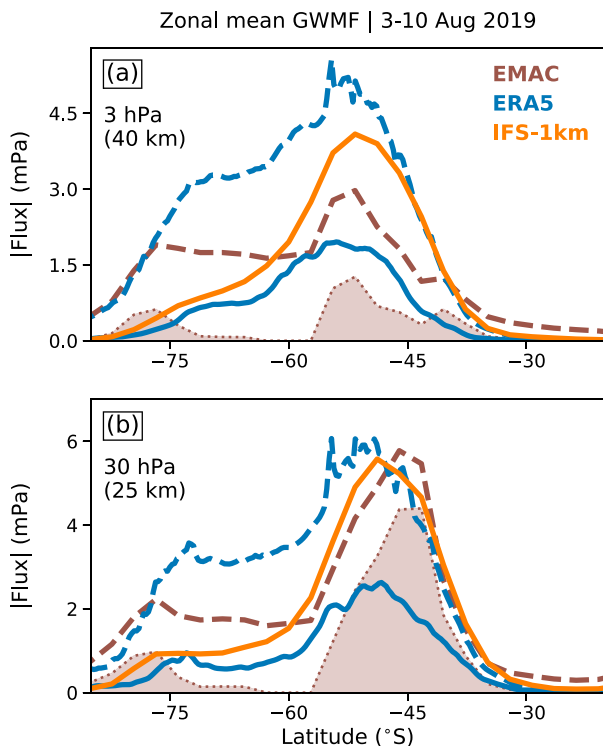
FIG. 4. Zonal-mean zonal GWMF at (a) 3 hPa (40 km) in the upper stratosphere and (b) 30 hPa (25 km) in the middle stratosphere during the period 3–10 Aug 2019 for the three models. The solid orange and blue curves show the total resolved fluxes (orographic plus nonorographic) in IFS and ERA5, respectively. The dashed blue curve shows the total resolved + parameterized (orographic + nonorographic) fluxes in ERA5, and the dashed brown curve shows the total orographic plus nonorographic parameterized fluxes in EMAC. The brown-shaded region enclosed within the dotted brown curve shows the orographic component of the parameterized flux in EMAC.

FIG. 5. As in Fig. 4, but with fluxes averaged only over longitudes 76°–20°W, i.e., around the Andes.

Similar features are noted when considering fluxes only in the vicinity of the Andes, i.e., averaging only over longitudes 76°–20°W (Fig. 5). This longitudinal span was chosen to completely encompass the gravity wave packet excited over the Andes that propagated laterally over to the Southern Ocean. At 25 km, almost all the flux in EMAC over the Andes is from orographic sources and is stronger than resolved fluxes in IFS-1km. More than two-thirds of the flux in EMAC dissipates between 25 and 40 km height; this is not the case for IFS-1km and ERA5. Poleward of 55°–60°S, the parameterized fluxes in both EMAC and ERA5 are a factor of 2–3 stronger than the resolved fluxes in IFS-1km. Between 55° and 60°S, both the resolved and parameterized fluxes rapidly decrease. This decrease is much more notable for the parameterized fluxes than it is for the resolved fluxes because of strictly zero contribution from the orographic parameterization in this region. This continuity in the forcing structure of resolved fluxes which is absent for the parameterized fluxes is due to contributions from lateral propagation of the resolved waves toward higher latitudes.
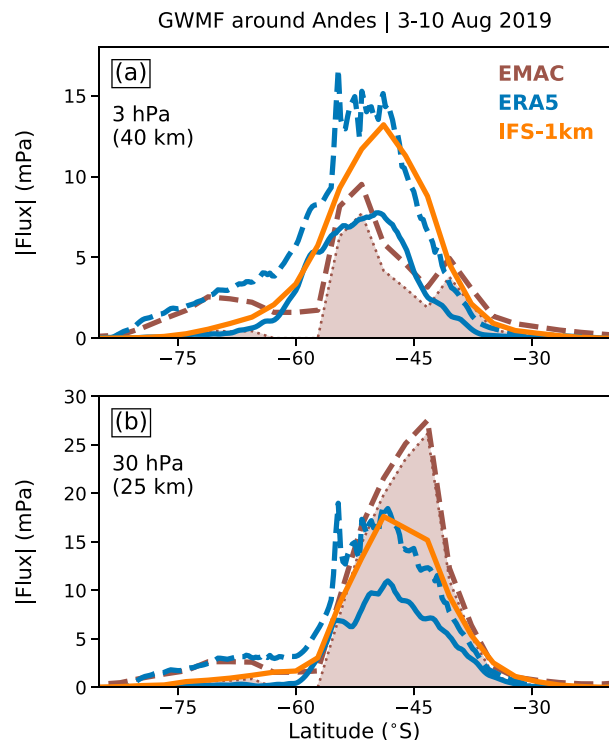
The zonal mean demonstrates that increasing the horizontal resolution (from ERA5 to IFS-1km) leads to improvements in resolved fluxes at all latitudes (Fig. 4). The regional average, however, suggests that, around the Andes, where the fluxes are predominantly orographic, the improvements are mostly seen equatorward of 50°S (Fig. 5). This indicates that the improved fluxes at all latitudes in the zonal mean most likely have nonorographic origins away from the Andes.

We show the VMFC, i.e., the mean zonal forcing from the vertical convergence of the GWMF, around the Andes in Fig. 6. In all three products, the dissipation in the stratosphere almost wholly occurs equatorward of 60°S. The resolved dissipation in both ERA5 and IFS-1km is very similar, reaching up to -40 m s^{-1} day^{-1} near the stratopause. The parameterized forcing in EMAC (Fig. 6c), however, has a strikingly different structure and begins much lower in the stratosphere (around 6 hPa) compared to the resolved VMFC (Figs. 6a,b). This highlights deficiencies in the parameterizations. As shown by Eichinger et al. (2024), lateral GW propagation can lead to higher GWMF deposition, leading to higher and stronger GW drag in GCMs and potentially reducing this parameterized versus resolved forcing difference.

We compare VMFC only below 1 hPa as dissipation above this level in IFS-1km and ERA5 is (partly) artificially induced by the diffusive sponge. As explained in section 2, for any given scale, the sponge dissipation in ERA5 is stronger than in IFS-1km, producing comparable VMFC in ERA5 around 1 hPa despite weaker resolved fluxes. Due to enhanced dissipation, all the resolved fluxes in ERA5 are completely dissipated by
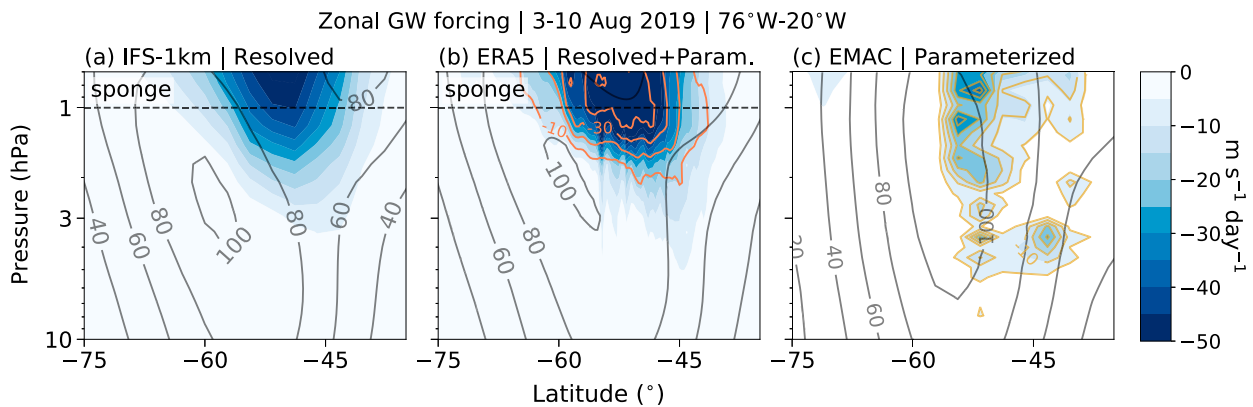
FIG. 6. Zonal GW forcing (color) and zonal winds (black curves) selectively averaged over longitudes of 76°–20°W around the Andes, during the period 3–10 Aug 2019 in (a) IFS-1km, (b) ERA5, and (c) EMAC. In (b), the resolved forcing is shown using orange curves, and the resolved plus parameterized forcing is shown in color. In (c), the orographic forcing from parameterization in EMAC is shown using yellow curves, and the total forcing is shown in color. The contour intervals for the orange and yellow curves in (b) and (c) are -10 and -5 m s^{-1} day^{-1}, respectively.

0.1 hPa while the dissipation in IFS happens more gradually and continues to near the model top at 0.01 hPa (not shown).

Both the parameterization design and treatment of the numerical damping at upper levels in high-top models continue to be a source of uncertainty for GW representation in the mesosphere where GW dissipation has leading-order effects. However, differences in the background flow and stratification profile among IFS-1km, ERA5, and EMAC may also contribute to the noted differences in GW momentum flux and forcing. For instance, EMAC's zonal-mean zonal winds lack the vertical tilt in the polar vortex, which is present in ERA5 and IFS-1km (Fig. 6). The wind differences between IFS and ERA5 are quite minor when compared to EMAC, so hereon, we focus only on resolved GWs in IFS-1km and ERA5, and validate them with high-fidelity observations from the CORAL lidar.

b. Comparing GWs in IFS-1km and ERA5 with observations from CORAL

1) GW TEMPERATURE PERTURBATIONS

Comparing the individual wave structure between models, reanalysis, and observations helps validate GW representation in high-resolution models, and allows addressing two key questions: (i) To what extent do differences in individual wave structure (amplitude, orientation, spectrum, etc.) dictate the differences noted in the GW fluxes, and (ii) can a 1.4 km hydrostatic model comprehensively resolve the GWs that constitute the complex wave envelope excited over the Andes?

We filter the temperature profile measured by CORAL to retrieve GWs over Rio Grande, and compare them with the resolved waves in IFS-1km and ERA5 during the first two weeks of August 2019. As detailed in section 3, the three datasets were treated identically by applying a fifth-order Butterworth filter to obtain GW-associated temperature perturbations from the full temperature profile. The retrieved GWs are shown in Fig. 7. Well-defined positive and negative alternating stripes represent strong GW activity in the middle atmosphere for all the early August nights when the lidar was active (Fig. 7a). The

GWs intensify with height and reach amplitudes of up to 25 K in the mesosphere. For days outside CORAL's coverage, perturbations from ERA5 (Fig. 7b) reveal continuous GW activity over Rio Grande throughout the 10-day period of 2–11 August 2019.

For all nights when CORAL was active, strongly in-phase positive (warm) and negative (cold) wave-induced anomalies are noted between CORAL and ERA5, as can be seen by comparing the colors with black contours in Fig. 7a. The amplitude, however, differs. The waves in IFS-1km (Fig. 7c), too, match well with ERA5 and CORAL, from 2 to 9 August. Especially for 2–6 August, IFS and ERA5 resolve GWs with strongly similar phase evolution and negligible phase shift. The phase differences become explicit only after 7 August. Eventually, beyond 9 August, key differences manifest as ERA5 shows GWs with amplitude above 20 K, but IFS shows only weak perturbations. Beyond this 10-day period, the local dynamics in the free-running IFS diverge from ERA5. In fact, around 13 August, IFS spontaneously generates strong GWs, which are not seen in ERA5 (Fig. 7c).

The vertical wave profiles for a selected day, 5 August, are shown in Fig. 8a. For this day, the GWs in the three products have a remarkably similar structure. In terms of amplitude, observed waves in CORAL are the strongest, and the waves in ERA5 the weakest. The modeled waves in both IFS and ERA5 are also visibly smoother than the observed waves in CORAL, despite identical filtering. The smoothness is predominantly due to the limited vertical resolution of the models, which is much lower than CORAL's operational resolution of 0.1 km. Their resolutions might be comparable in the lower stratosphere, but are considerably coarser than CORAL's in and above the upper stratosphere. As a result, the finer-scale variations in CORAL contain noticeable contributions from vertical scales which are smaller than those resolved in ERA5 and IFS.

The difference in wave amplitude is illustrated more clearly in Fig. 8b. Here, the sinusoidal signals have been converted to amplitudes through vertical integration. Throughout the vertical,

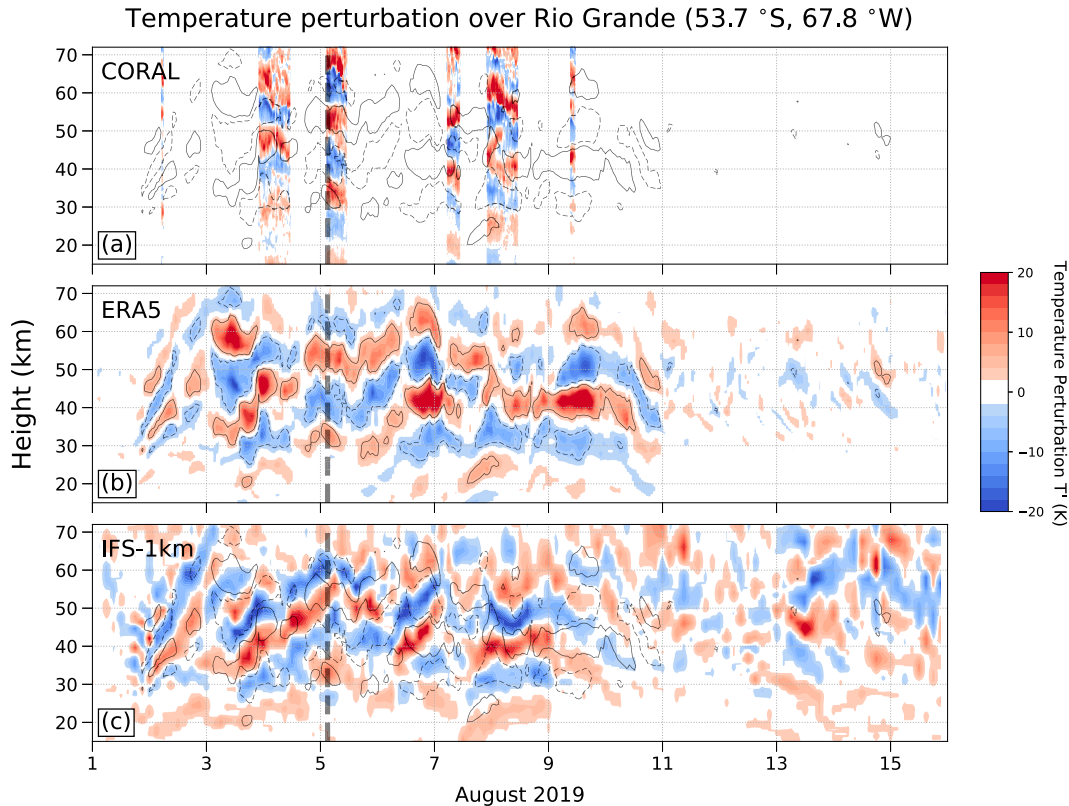## Temperature perturbation over Rio Grande (53.7 °S, 67.8 °W)

FIG. 7. Vertical profiles of GW temperature perturbations (color) during 1–15 Aug 2019 over Rio Grande, Argentina (53.7°S, 67.8°W) from (a) ground-based CORAL stationed at Rio Grande, (b) ERA5, and (c) ECMWF's IFS-1km free-running model. The black curves in (a)–(c) show the ±5 K temperature anomaly contours from ERA5, and are overlaid for reference. The dashed vertical bars in (a)–(c) mark the time 0300 UTC 5 Aug 2019 for which the horizontal profile of the GWs for ERA5 and IFS-1km are shown in Fig. 9.

ERA5 yields the weakest wave amplitudes, while CORAL yields the strongest amplitudes. For all products, the amplitudes increase with height between 25 and 45 km, due to a reduction in background density. Above 45 km, the amplitude of the modeled waves in ERA5 and IFS-1km weakens. In contrast, wave amplitudes in CORAL amplify from 10 to 15 K near 45 km, following a reduction in the background static stability around the stratopause (1 hPa). In ERA5 and IFS-1km, any such amplification is overshadowed by attenuation by the numerical sponge introduced at 1 hPa (48 km). Therefore, due to a combined effect of limited vertical resolution and a numerical sponge above 1 hPa, IFS-1km, even at an unprecedented 1.4 km horizontal grid spacing, generates GWs that are weaker than those observed by CORAL.

A comparison of horizontal profiles of the resolved GWs in the upper stratosphere for the same day, 0000 UTC 5 August is shown in Fig. 9. The horizontal structures between ERA5 and IFS-1km have nearly identical dominant horizontal wavelengths and orientation, but with a slight phase shift over the Southern Ocean (Figs. 9a,b). Here, notable contributions from the finer-scale waves in IFS-1km are apparent throughout the domain, capturing the improvements obtained by a resolution increase. These improvements are illustrated by comparing the perturbations at

53.7°S (Fig. 9c). Comparing waves in the 1 km model with a 9 km IFS run, Polichtchouk et al. (2023) found that the zonal-mean contribution from such newly resolved waves, i.e., waves with horizontal wavelengths less than 100 km, could form up to 50% of the total GW forcing.

To better understand the contribution from various scales, we compute the power spectrum for August 2019 and compare it with the 20-yr-averaged climatological power spectrum (Fig. 10a). The averaged power spectrum was obtained by selectively averaging over days with moderate-to-strong GW activity ($\beta \geq 10$ K^2). The spectrum reveals sufficient power at 45 km at scales from 300 to 1200 km, and maximum power predominantly due to Andean mountain waves within zonal wavelengths from 500 to 800 km (Fig. 10a; bold black curve). Interestingly, the August 2019 power spectrum for ERA5 (dashed red) is quite similar to the 20-yr mean. The spectrum tapers sharply around $\lambda_x \leq 250$ km due to gridscale hyperdiffusion and other numerical choices within a dynamical core which reduce the effective horizontal resolution of ERA5 to ≈200–250 km. The 15-day mean power spectrum for IFS is similar to ERA5 at large scales but also contains considerably more power at scales much finer than 300 km (bold orange). The (scaled) spectrum peaks around 600 km, which is slightly
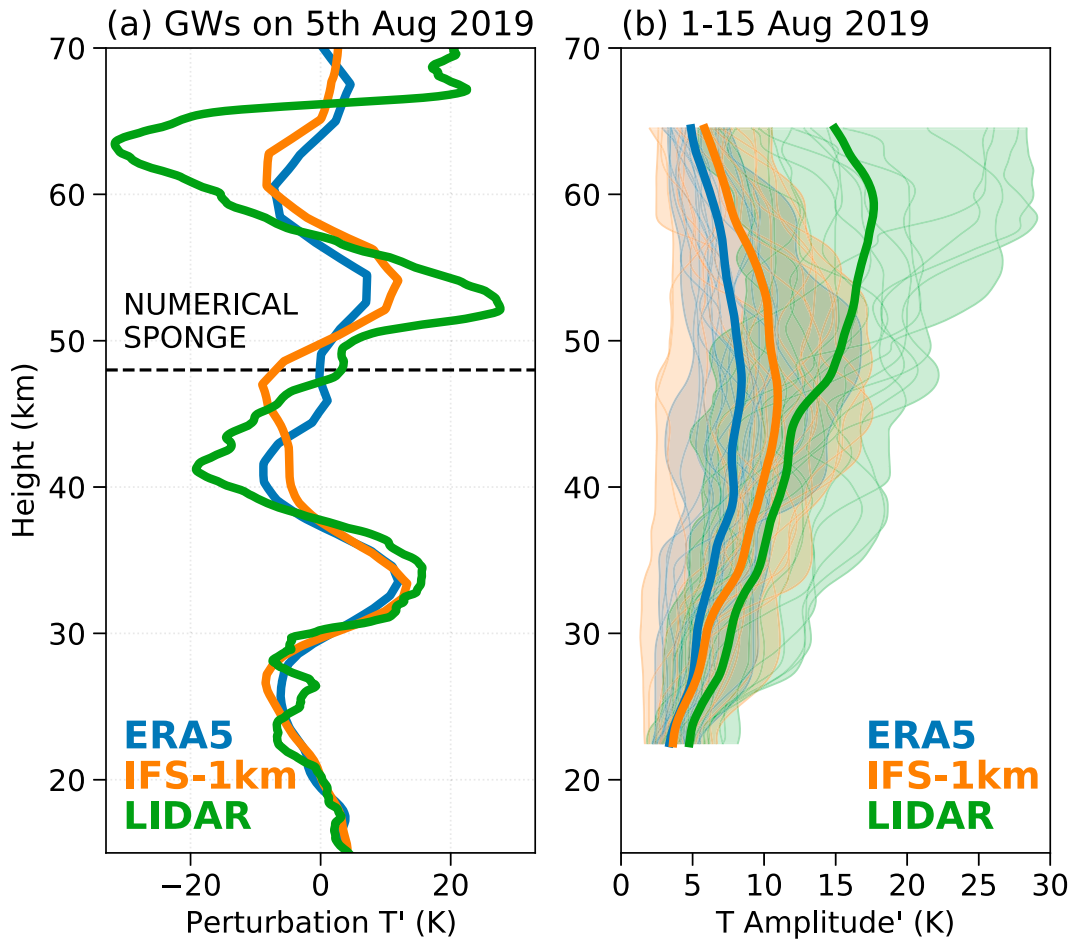
FIG. 8. (a) Vertical profiles of temperature perturbations over Rio Grande in CORAL (green), IFS (orange), and ERA5 (blue), at 0300 UTC 5 Aug 2019. The dashed black horizontal bar at 48 km marks the introduction of the mesospheric sponge. (b) Gravity wave amplitudes in the three models, obtained by vertically averaging the temperature magnitude over a centered 15 km vertical window. The bold curves show the time mean amplitude, while the individual thin curves enclosed within the respective shaded regions show the amplitudes at all the individual snapshots during 1–15 Aug 2019 for which the data are available.

lower than ERA5 which peaks at 650 km, indicating that the waves resolved in IFS, on average, have slightly shorter wavelengths than the waves resolved in ERA5.

The temporal evolution of the IFS spectrum for 1–15 August 2019 in the upper stratosphere (42 km) is shown in Fig. 10b. The evolution is similar to that for ERA5 (Fig. S8) and shows the intermittent increase in power due to excited GWs. Most prominently, the spectrum shows the transfer of power from shorter to longer wavelengths, easily seen on 4, 6, and 8 August. The transfer suggests rotation and possible refraction experienced by the upward-propagating GWs as they propagate through a strongly sheared flow.

2) ESTIMATING GW MOMENTUM FLUX ACROSS CORAL, IFS-1KM, AND ERA5

A plausible validation of the vertical and horizontal wave profiles motivates a comparison of GWMF across the datasets. We perform a set of three comparisons:

C1. *Local proxies*: First, we compute the temperature-based momentum flux proxy in CORAL, IFS-1km, and ERA5, using the temperature profile *directly over* Rio Grande (Fig. 11a), i.e., using temperature perturbations only over a fixed latitude–longitude (since CORAL measures temperature only above Rio Grande). This allows a local, like-for-like comparison among observations, models, and reanalysis.

C2. *Coarse-grained conservatively interpolated proxies*: Computing fluxes using C1 includes wave transience effects associated with different wave phases of a passing wave, yielding a noisier flux evolution (Ern et al. 2004; Procházková et al. 2023). This can be resolved by averaging the flux over single or multiple wave cycles. Therefore, we compute a second temperature-based proxy for all the points on the ERA5 and IFS-1km native grid in a 1000 km × 1000 km box around Rio Grande, then conservatively coarse graining them onto a T21 Gaussian grid, and then (linearly) interpolate the conservative proxies onto Rio Grande's latitude and longitude (Fig. 11b).
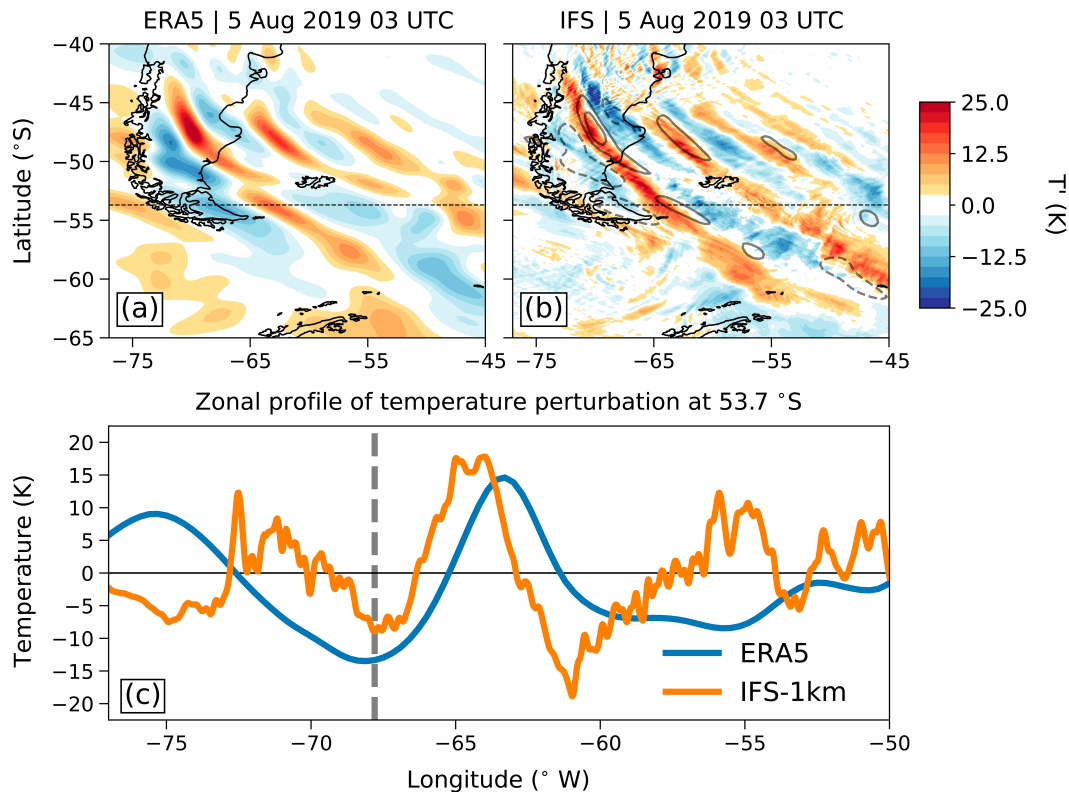
FIG. 9. (a),(b) Horizontal structure of the resolved GW perturbations at 1.5 hPa (45 km) height, at 0300 UTC 5 Aug 2019 in ERA5 and IFS, respectively. The temperature anomalies were retrieved from the full temperature field using a high-pass filter. (c) Comparing temperature perturbation profiles in (a) and (b) at Rio Grande's latitude. The dotted black lines in (a) and (b) and the dashed black line in (c) respectively mark Rio Grande's latitude (53.7°S) and longitude (67.8°W).

C3. *Actual momentum flux*: Third, we compute the GWMF from models directly and compare them with the temperature-based proxy from CORAL (Fig. 11c). Here, for consistency, we first compute the covariances on the native ERA5 and IFS-1km grid, then conservatively coarse grain them onto a T21 Gaussian grid, and finally (linearly) interpolate them onto Rio Grande's latitude and longitude. The coarse-graining details to compute C2 and C3 are provided in section 3c.

Temperature-based C1 demonstrates similar variability for both IFS and ERA5 (Fig. 11a). For instance, they both intensify around 0000 UTC 2, 1200 UTC 3, and 1800 UTC 6 August (marked with blue and orange triangles). The proxy C1 from IFS is, in general, stronger than the proxy from ERA5. Moreover, C1 from CORAL on 5 and 7 August is stronger than the proxy from IFS. This is consistent with the wave amplitudes of Fig. 8b. Throughout, the proxy from CORAL agrees more strongly with ERA5 than it does with IFS.

The proxies, C1, exhibit some notable differences as well. For instance, the proxy from IFS intensifies around 1200 UTC 4 August but CORAL shows an intensification more than 12 h later, while ERA5 does not show any intensification at all. A similar pattern occurs around 0000 UTC 7 August.

Some of the disagreements in C1 proxies in Fig. 11a due to wave transience are eliminated after conservatively integrating

the proxies over single/multiple wave cycles (Fig. 11b). For example, conservative interpolation eliminates the difference in IFS-1km and ERA5 C1 proxies at 1200 UTC 4 August, and produces more similar C2 proxies across the three products. The modified proxy, C2, from IFS-1km is stronger than C2 from ERA5, with an exception on 3 August, where the C2 proxies in IFS-1km and ERA5 are nearly identical.

The wind-based fluxes C3 (Fig. 11c) match the evolution of the temperature-based proxy, C2 (Fig. 11b), quite well. The key difference between C2 and C3 is in the magnitude. For most dates, C2 and C3 have similar magnitudes, with C3 being slightly stronger. However, most notably on 5, 6, and 9 August for ERA5, the estimated fluxes from C2 are stronger than C3.

c. Sensitivity of the flux estimates to the formulation used

The estimated fluxes can be sensitive to the formulation used. While the two formulations, following appropriate averaging, are largely similar in magnitude and evolution, they also disagree in various instances. How systematic, then, are these similarities/differences? To disentangle this sensitivity from a potential sampling bias, we leverage the multidecadal coverage of ERA5 to extend the comparison to a 20-yr period from 2000 to 2019. Essentially, we compute and compare GWMF using the two
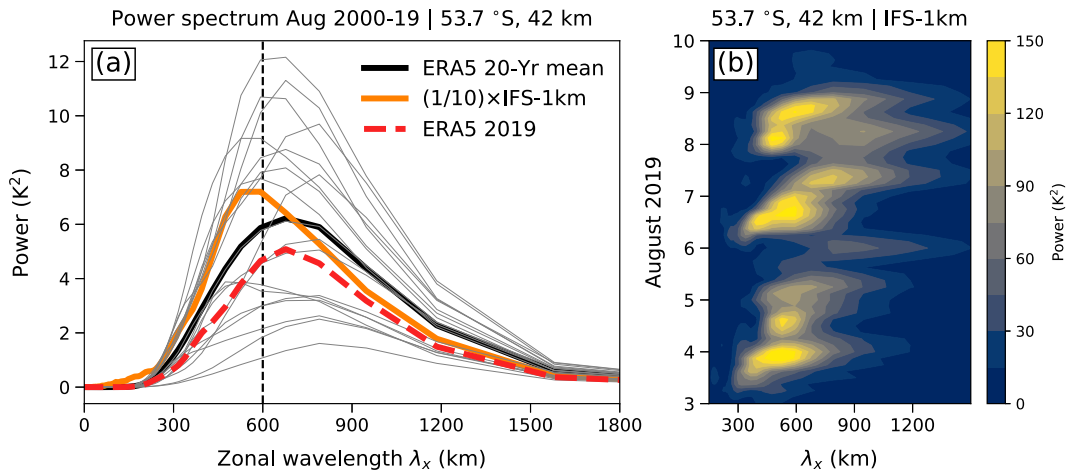
FIG. 10. (a) The 2000–19-averaged zonal power spectrum of the temperature from ERA5 between 90°W and 0° at 1.5 hPa (45 km) height for August, for days with moderate-to-strong GW activity in the upper stratosphere ($\beta \gtrsim 10\,\mathrm{K}^2$). The bold orange curve shows (scaled) power spectrum from IFS-1km averaged over 3–10 Aug, the bold black curve shows the 20-yr mean spectrum, the dashed red curve shows the mean spectrum for August 2019, and the light gray lines show the power spectrum for the individual years. The dashed vertical bar shows the wavelength associated with peak power from IFS. (b) The power spectrum at 1.5 hPa (42 km) for 3–9 Aug 2019, as seen in IFS.

formulations for the austral winter months of June, July, and August, for days with moderate-to-strong GW activity ($\beta \gtrsim 10\,\mathrm{K}^2$) in the upper stratosphere. To compute the temperature-based proxy over the 20-yr period, we compute the dominant horizontal and vertical wavelength separately for each day using a 1D zonal Fourier transform around Rio Grande and a 1D Fourier transform in the vertical.

The comparison between the two formulations over the 20-yr period is presented in Fig. 12. The scatterplot indicates that, in a climatological sense, the fluxes obtained from the two formulations are quite similar, and line up close to the one-to-one line, more so for strong GWs (Fig. 12a). A best-fit slope of $\sim$0.78 $\pm$ 0.15 suggests that, on average, the temperature-based fluxes have a 25% larger magnitude than the wind-based fluxes. Both moderate and strong GW cases have sufficient scatter points above the one-to-one line as well, indicating that the wind covariances estimates can indeed sometimes be stronger than estimates from temperature-based proxies, explaining our findings in Fig. 11. Therefore, in a statistical sense, one formulation does not provide an upper bound on the other.

The temperature-based proxy relies only on the local temperature profile and attributes all the perturbations to a dominant horizontal (λ_x) and vertical (λ_z) wavelength. This severe approximation treats any complex wave packet as a monochromatic wave with a given aspect ratio (λ_z / λ_x). While this formulation allows an estimation of fluxes from in situ and remote sensing temperature observations in lieu of sufficient wind records, it does so by introducing two free parameters, λ_x and λ_z. The estimation of these free parameters can potentially introduce considerable uncertainty in the flux estimation. Estimates of these wavelengths from raw temperature data could be sensitive to the methodology used. Moreover, there is no direct way to compute the horizontal wavelengths using vertical profilers such as CORAL.

Uncertainty in the estimates of the "effective" horizontal and vertical wavelengths can be highlighted using the wave aspect ratio λ_z / λ_x, which is required to estimate the momentum fluxes from temperature perturbations. Assuming the wave aspect ratio as an unknown free parameter (which itself contains two free parameters) and equating the fluxes from the two formulations, we compute the *inferred* distribution of the aspect ratio that would yield equal flux estimates. The inferred distribution (Fig. 12b) suggests that computing wavelengths using Fourier transforms from ERA5 may lead to a systematic overestimation of the wave aspect ratio leading to higher temperature-based proxies. The mean aspect ratio from the inferred distribution is about a factor 1.25 lower than the mean aspect ratio extracted from ERA5.

Effectively, uncertainty in wavelength estimates can introduce notable errors when estimating fluxes exclusively using temperature-based proxies. Around the Andes, it can lead to a potential overestimation of true momentum fluxes.

5. Discussion and conclusions

The first 10 days of August 2019 were a period of intense GW activity in the extratropical middle atmosphere predominantly due to large-amplitude mountain waves excited over the Andes. We assessed GW representation and fluxes in the austral stratosphere during the first two weeks of August 2019 by blending a 1.4 km global forecast model (ECMWF IFS) and observations from a ground-based Rayleigh lidar (CORAL), with modern reanalysis (ERA5) and a coarse-resolution chemistry–climate model (EMAC). A comprehensive analysis of this kind, one which spans the whole range of GW resolutions—from fully resolved to fully parameterized, and from local representation to regional
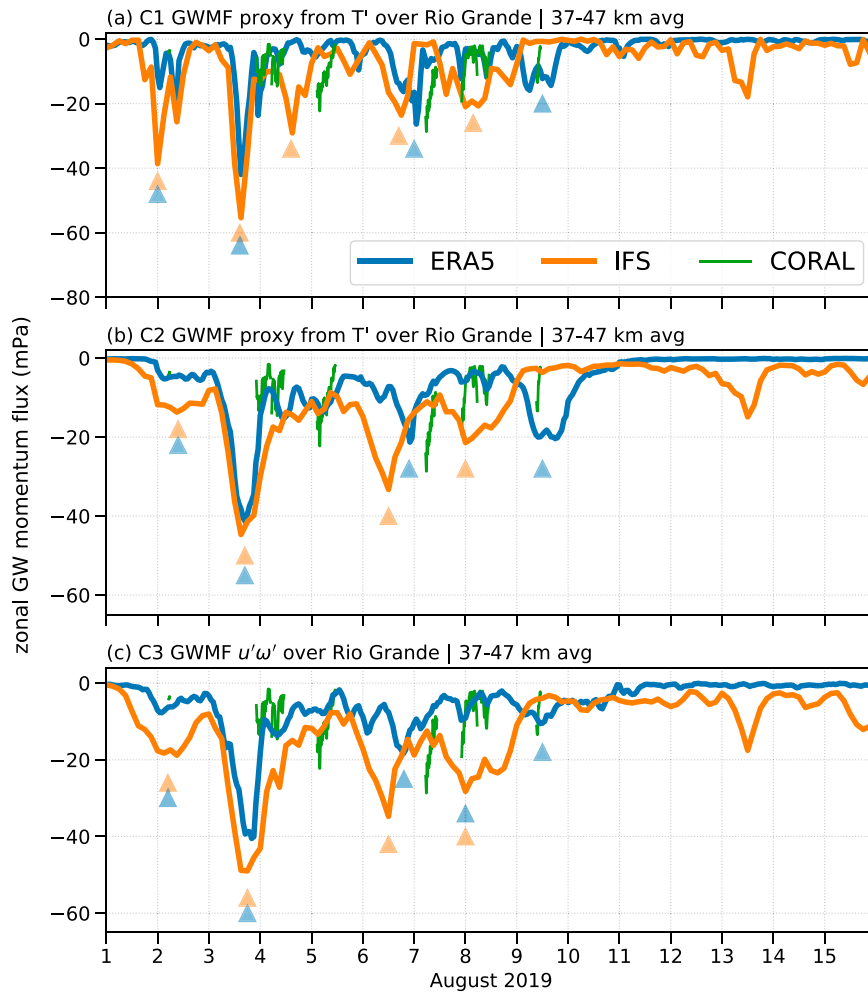
FIG. 11. A comparison of zonal GWMF and its proxy (in mPa) in CORAL (green), IFS (orange), and ERA5 (blue) during 1–15 Aug 2019 averaged between 37 and 47 km height in the upper stratosphere. (a) GWMF proxy calculated using C1 from temperature profiles over Rio Grande. (b) GWMF proxy from IFS-1km and ERA5 calculated using C2 from the temperature profiles with subsequent coarse graining to T21 Gaussian grid and interpolation to Rio Grande. (c) Directly computed GWMF over Rio Grande from IFS-1km and ERA5 calculated using C3. The orange and blue triangles mark peaks in IFS and ERA5 fluxes/proxies, respectively. The temperature-based proxy from CORAL (green) is identical in all three subplots.

and global forcing—has not been conducted within a single study before. Employing the broad range of datasets allowed the validation of a kilometer-scale model using ground-based observations, and facilitated a critical comparison of both resolved and parameterized GW fluxes in the stratosphere. It also allowed the evaluation of different flux approximations based on linear-wave theory for middle atmospheric GW analysis.

IFS, ERA5, and EMAC had similar wind profiles, especially for the first 10 days in August, because the tropospheric and lower stratospheric winds in EMAC were nudged to ERA5, and the free-running IFS-1km was initialized on 1 August 2019 with operational IFS conditions. This provided a time window to meaningfully assess the stratospheric GWMFs in ERA5 and EMAC, against the fully resolved "model truth" in IFS, both in the zonal mean and regionally around the Andes.

a. How do the resolved fluxes in the ~1 km "model truth" versus ERA5 compare?

We found a strong qualitative agreement in *resolved* stratospheric GWMFs between ERA5 and IFS-1km. A factor of ~20 finer horizontal resolution in IFS-1km produced, on average, a factor of 2–2.5 stronger GWMF in the zonal mean and over the Andes during 3–10 August 2019. In other words, ERA5 severely underestimates and, on average, resolves only ~40% of the "model truth," highlighting significant contributions from unresolved waves with wavelengths shorter than 200 km.
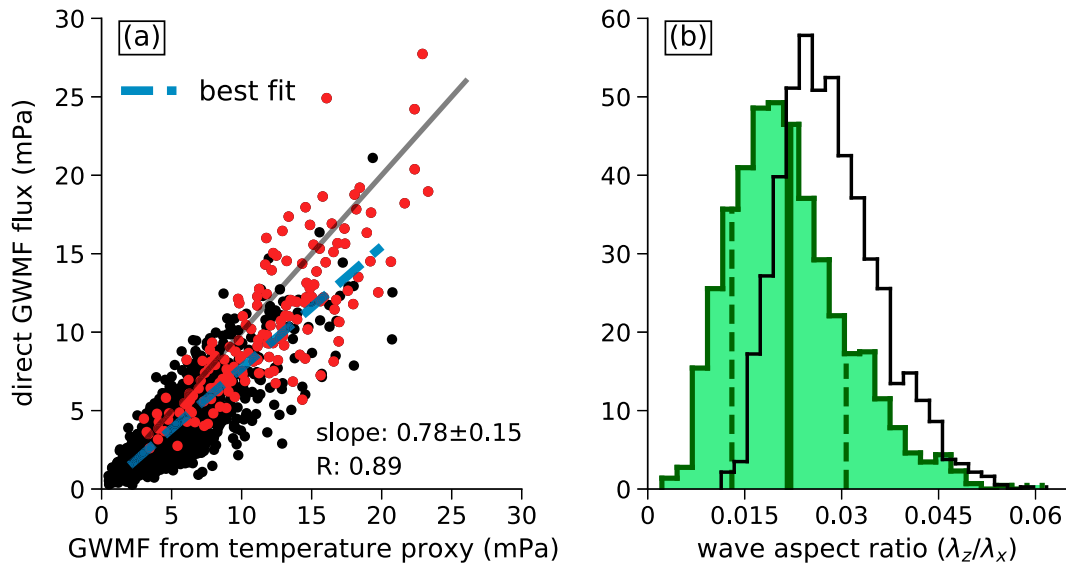
FIG. 12. (a) A scatterplot illustrating a comparison between the temperature-based proxies in ERA5 vs the fluxes computed directly, for JJA 2000–19. Only the fluxes during austral winter days when moderate-to-strong GWs ($\beta \geq 10$ K^2) were detected in the upper stratosphere are considered. The black and red points represent moderate ($10 \leq \beta \leq 100$ K^2) and strong GWs ($\beta \geq 100$ K^2), respectively. The one-to-one line is shown in black, and the best-fit line is in dashed blue. Slope uncertainty in (a) corresponds to 95% confidence. (b) The black envelope shows the distribution of the aspect ratio λ_z/λ_x, computed directly from ERA5 using Fourier transforms and used to compute the scatterplots in (a). The green histogram shows the distribution inferred by equating the fluxes from the two formulations [Eqs. (3) and (4)]. The solid and dashed dark green vertical bars show the mean and one sigma standard deviation, respectively.

b. How well do the parameterized fluxes in coarser models compare with the ~1 km "model truth"?

The comparison found prominent contrasts between the *resolved* and *parameterized* GWMFs. The parameterization identifies mountain wave generation over the Andes and generates strong orographic fluxes. However, the resolved and parameterized flux magnitudes notably differ at all latitudes in the upper stratosphere. In fact, poleward of 60°S, the parameterizations in ERA5 and EMAC generate excessively strong (majorly nonorographic) fluxes, which are 3 times higher than the "model truth."

The parameterized versus resolved fluxes are dissipated very differently among the models most likely because the orographic parameterizations completely ignore lateral wave propagation and wave transience. Generally speaking, all operational orographic and nonorographic GW parameterizations neglect these effects. These simplifications still provide somewhat competent GWMF over the Andes following some tuning, but they fail to accurately represent the spatiotemporal evolution of the flux (like propagating from the Andes to over the ocean) resulting in biases like little-to-no mountain wave–induced variability over the Southern Ocean (Fig. 13, brown curve versus other curves). Recent improvements in GW parameterizations have been proposed through the development of new schemes that include multichromaticity (van Niekerk and Vosper 2021), and lateral propagation effects either through momentum flux redistribution (Eichinger et al. 2024) or using WKB-based ray tracing schemes (Amemiya and Sato 2016; Voelker et al. 2023).

c. Model validation of the IFS-1km using CORAL and ERA5

Similar background flow also permitted a comparison of horizontal profiles of the resolved GWs between IFS and ERA5, and a validation of their vertical profiles with CORAL measurements around Rio Grande. The resolved waves in IFS-1km and ERA5 bear similar vertical and horizontal phase profiles of temperature perturbations for the first 9–10 days in August. The vertical phase profiles also agreed well with CORAL for all the nights it was active. The datasets most notably disagreed on the wave amplitudes: GWs in ERA5 were weaker than in IFS, and GWs in IFS, even at a ~1 km resolution, were weaker than in CORAL. The model wave amplitudes also remarkably weaken above 1 hPa whereas CORAL amplitudes steadily increase with height. This rapid attenuation in models is due to the sponge damping the modeled GWs. Waves in CORAL also contained finer-scale vertical variations not seen in temperature perturbations in ERA5 and IFS, most likely due to limited vertical resolution.

The artificial sponge damping in the mesosphere and above prohibits us from fully leveraging the improvements achieved by the horizontal-resolution increase in IFS-1km. At present, the only global products without a mesospheric sponge are the select reanalyses with model tops higher than 100 km (e.g., Eckermann et al. 2018). Due to sponge effects, the upper atmosphere region in the models combines the physical dissipation of wave energy from the remaining resolved GWs with their progressively stronger numerical damping toward
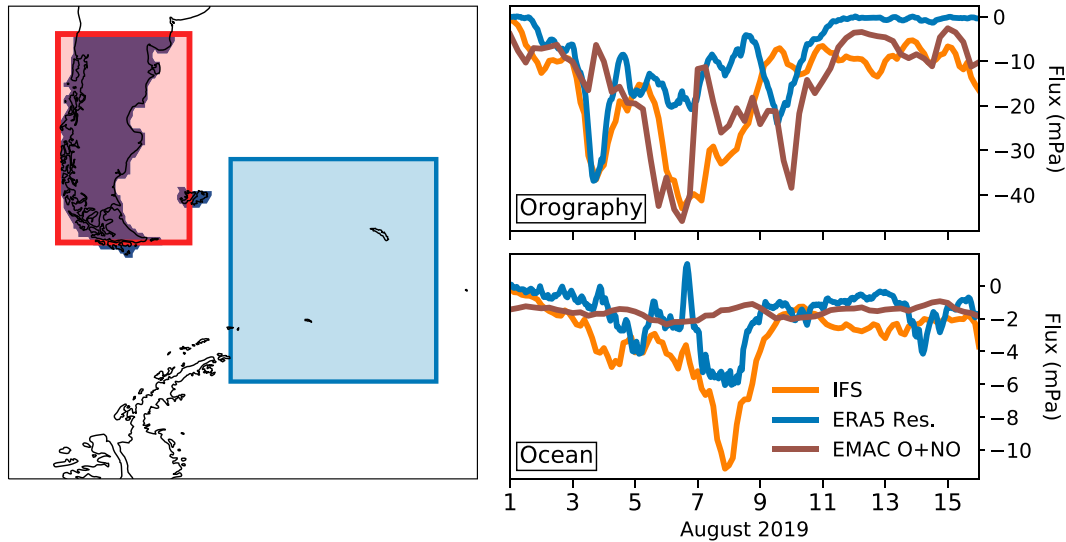
FIG. 13. Comparing momentum flux variability in IFS (orange curves), ERA5 (blue curves), and EMAC (brown curves), over the continental orography and the Southern Ocean, at 10 hPa (30 km) height. (b) Integrated momentum flux over the South American subcontinent, i.e., the red box in (a). (c) Integrated momentum flux over the Southern Ocean, i.e., the blue box in (a).

the model top. Currently, skillful tuning of numerics and GW parameterizations in this region is used to obtaining reliable predictions of the tropospheric and stratospheric circulations. For the same reason GW forcing in different reanalyses peaks at different heights in Fig. 1. Potential improvements in GW parameterizations could altogether eliminate the need for such tuning and artificial modulation of the middle-atmosphere circulation by it.

d. GW momentum flux estimates across models, observations, and reanalysis

The validation of temperature perturbations in an ultra-high-resolution global model (and reanalysis) using CORAL strengthens confidence in their ability to simulate GWs in both constrained and free-running configurations. Yet, for a truthful representation of the stratospheric dynamics and circulation, it is imperative that the models also agree on the resolved GWMF. Models can diagnose the vertical velocity and provide global coverage of horizontal winds but most remote sensing instruments (including CORAL) only measure temperature. To bridge this gap, we used the validated GW temperature profiles and model winds to estimate GWMFs around Rio Grande using wind covariance and a temperature-based proxy of the momentum flux. Both quantities were derived as an approximation to the vertical flux of horizontal pseudomomentum flux from linear wave theory: one using wind covariance (i.e., the actual momentum flux computable from models), and the other using temperature variance (used for observational estimates) obtained by expressing the wind covariance in terms of temperature and applying monochromaticity and midfrequency approximations.

Comparing the two GWMF approximations revealed that the flux estimates can be sensitive to the formulation itself.

The two formulations provided similar flux evolution for most times across datasets, providing confidence in our flux estimates. However, on certain occasions, their estimates were quite different. For instance, the temperature-proxy estimates around 9 August were a factor of 4 stronger than the wind-covariance estimates in ERA5. Also, temperature-proxy estimates suggest very similar fluxes in ERA5 and IFS on 3 August, but the wind covariances yield 20%–25% stronger fluxes in IFS.

The fluxes in IFS *directly over* Rio Grande are not substantially stronger than ERA5, despite different model resolutions. We understand this by connecting the local flux estimates above Rio Grande with the regional flux estimates (Fig. 5), which indicate that most improvements in GWMF in IFS in the upper stratosphere are obtained equatorward of 50°S and are obscured in local estimates because the lidar profiles temperature only over Rio Grande (53.7°S). This highlights the need for a more extensive lidar coverage in the horizontal. In fact, the regional fluxes at 53.7°S are only slightly stronger than ERA5. The power spectrum at 53.7°S indicates that a majority of GW energy here is contained within long waves, i.e., wavelengths 400 km and higher. These scales are fairly well resolved in ERA5. Poleward of 50°S, the improvements in resolved fluxes in IFS are visibly higher in the zonal mean than they are around the Andes. Thus, further from the Andes, the contribution from shorter wavelengths becomes more prominent due to GWs from other sources. These sources could include nonorographic waves generated over the Southern Ocean or possibly even mountain waves generated over small islands in the Southern Ocean (Hendricks et al. 2014; Hindley et al. 2021). These waves can contribute significantly to the zonal momentum budget but are barely resolved even in a 9 km IFS run (Polichtchouk et al. 2023).

We tested for possible systematic differences between proxies on climatological time scales and found that the temperature-

based proxy tends to produce flux estimates which are, on average, 25% higher in magnitude than those from wind covariance. These differences, including the differences for the August 2019 period, can be explained by a combination of two key factors: (i) wavelength estimation uncertainty, and (ii) approximation uncertainty. (i) refers to the uncertainty involved in the estimation of characteristic horizontal and vertical wavelengths. The temperature-based proxy assumes the monochromaticity of complex wave packets—attributing all perturbations to a single, characteristic horizontal wavelength, estimating which can be sensitive to the method and spatial averaging and filtering used. Undeniably, local transforms such as the S3D transform (Chen et al. 2022) could possibly provide better wavelength estimates. (ii) refers to uncertainty due to midfrequency approximations, i.e., the wind-based flux is obtained by approximating $(1 - f^2/\hat{\omega}^2) \approx 1$ in Eq. (2). From a transformed Eulerian mean (TEM) perspective, this is equivalent to ignoring the meridional heat flux by gravity waves. On the other hand, the temperature-based proxy is obtained by first expressing the wind covariance in terms of temperature perturbations, and then assuming wave monochromaticity and using midfrequency approximations of the dispersion relationship, $\hat{\omega}^2 = N^2 k_h^2/m^2$. [Here k_h is the total horizontal wavenumber. See Ern et al. (2004) for details.] Thus, both approximations to the pseudomomentum flux lead to different deviations.

Ern et al. (2017) illustrated the scope for further improvements in estimation of fluxes from satellite data using temperature-based proxies. Alternatively, recent studies have also proposed proxies based on wind and temperature covariances. For instance, Stephan et al. (2019) proposed the "wind and temperature quadratic" to estimate GWMF from gridded model data. Similar to the temperature-based proxy, however, this proxy is formulated exactly for monochromatic waves but fails to generalize well for multichromatic wave packets.

e. Conclusions

Temperature-based proxies are an invaluable tool to estimate fluxes from observations, both to understand the global GWMF distribution and to validate model parameterizations. However, as shown, caution is necessary when directly connecting them to validate models and model parameterizations. One fix could be more comprehensive satellite ground-based and remote sensing observations that measure the fine-scale horizontal and vertical winds over a sufficiently large horizontal and vertical expanse, eliminating the need to use temperature altogether. Doing so at a high temporal frequency would also allow proper estimation of the intrinsic wave frequency. Currently, the benefits accompanying the high spatial resolution of ground-based or airborne lidars are offset by their severely limited regional and temporal coverage. Augmenting the vertical information profiled by the lidar with horizontal information from other products to estimate fluxes can introduce inconsistencies and increase uncertainty.

A commensurate increase in the vertical resolution and improvements in upper-boundary damping in climate models and reanalysis, too, can open avenues for a more accurate GW analysis in the upper stratosphere and mesosphere by bringing models closer to observations.

Major differences between the resolved and parameterized GWMF call for an urgent improvement in GW parameterization. Arguably, some of the differences in GW dissipation we noted could be due to different model winds in the middle atmosphere or due to different sponge damping. Despite this, the overarching differences in GW dissipation profiles are clearly rooted in inadequately designed parameterizations. Guided by observational and high-resolution model benchmarks, new parameterizations must include lateral propagation, multichromaticity, wave–wave interactions, transient evolution of GWs, and account for zonal variations in GW sources. Particularly for mountain waves, better representation of subgrid-scale terrain heights and accounting for terrain anisotropy can guarantee improvements, as was shown by Kruse et al. (2022). The current practice to compensate for such lack of variability over orographic sources or the ocean is to overtune the nonorographic drag from zonally uniform GW sources to compensate for a lacking orographic drag in the upper atmosphere. While this might "fix" the GWMF and dissipation near topographical features, it potentially sets off dynamical imbalances in other regions (Polichtchouk et al. 2018, 2023). These imbalances when accumulated over seasonal time scales can ultimately generate a significant middle atmospheric circulation bias.

Summarizing, our analysis aimed to bridge linear-wave theory and GW observations with high-resolution models, reanalysis, and coarse-resolution climate models. Validation of an unprecedented 1.4 km global model using lidar observations demonstrates its capability to accurately simulate GWs without parameterizations and also informs that further improvements in modeling choices and grid resolution are necessary to match the observed GW amplitudes. Kruse et al. (2022) found that increasing model resolution improved validation, but GWs were still underrepresented or diffused even in 3 km state-of-the-art NWP models—our findings indicate that they are underrepresented even at a 1.4 km resolution. Validating modern reanalysis both with the "model truth" from IFS and the "observational truth" from lidar demonstrates its strength in partly simulating wave profiles. It also reveals that modern reanalysis still heavily relies on parameterizations to represent mesoscale variability in the stratosphere and above, and parameterizes more than half of the actual GW-induced forcing in the southern extratropics to represent the stratospheric and mesospheric circulation. Finally, testing the two formulations relying on midfrequency approximation, to estimate GWMF, demonstrates the need for better proxies, and the need for more comprehensive observations in order to translate GW observations to effective GW modeling.

Data availability statement. ERA5 data can be freely accessed from https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era5. Output from the CCMI-2022 EMAC model simulations is stored at the Centre for Environmental Data Analysis (CEDA) in the United Kingdom (https://blogs.reading.ac.uk/ccmi/ccmi-2022_archive/). Data on shorter time scales than stored at CEDA are available upon request. The raw output from the 1.4 km IFS model on the native grids amounts to a few hundred terabytes, so it is not possible to make all data available for more than a few steps. However, the postprocessed output used to make the figures will be retained and made available to those who request it.

REFERENCES

Albers, J. R., and T. Birner, 2014: Vortex preconditioning due to planetary and gravity waves prior to sudden stratospheric warmings. *J. Atmos. Sci.*, **71**, 4028–4054, https://doi.org/10.1175/JAS-D-14-0026.1.

Alexander, M. J., and H. Teitelbaum, 2011: Three-dimensional properties of Andes mountain waves observed by satellite: A case study. *J. Geophys. Res.*, **116**, D23110, https://doi.org/10.1029/2011JD016151.

——, 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. Atmos.*, **118**, 11589–11599, https://doi.org/10.1002/2013JD020526.

——, and Coauthors, 2010: Recent developments in gravity-wave effects in climate models and the global distribution of gravity-wave momentum flux from observations and models. *Quart. J. Roy. Meteor. Soc.*, **136**, 1103–1124, https://doi.org/10.1002/qj.637.

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

Banyard, T. P., C. J. Wright, N. P. Hindley, G. Halloran, I. Krisch, B. Kaifler, and L. Hoffmann, 2021: Atmospheric gravity waves in Aeolus wind lidar observations. *Geophys. Res. Lett.*, **48**, e2021GL092756, https://doi.org/10.1029/2021GL092756.

Baumgaertner, A. J. G., P. Jöckel, A. D. Aylward, and M. J. Harris, 2013: Simulation of particle precipitation effects on the atmosphere with the MESSy model system. *Climate and Weather of the Sun-Earth System (CAWSES): Highlights from a Priority Program*, F.-J. Lübken, Ed., Springer, 301–316, https://doi.org/10.1007/978-94-007-4348-9_17.

Becker, E., 2012: Dynamical control of the middle atmosphere. *Space Sci. Rev.*, **168**, 283–314, https://doi.org/10.1007/s11214-011-9841-5.

Chen, Q., and Coauthors, 2022: Satellite observations of gravity wave momentum flux in the mesosphere and lower thermosphere (MLT): Feasibility and requirements. *Atmos. Meas. Tech.*, **15**, 7071–7103, https://doi.org/10.5194/amt-15-7071-2022.

Corcos, M., A. Hertzog, R. Plougonven, and A. Podglajen, 2021: Observation of gravity waves at the tropical tropopause using superpressure balloons. *J. Geophys. Res. Atmos.*, **126**, e2021JD035165, https://doi.org/10.1029/2021JD035165.

Domeisen, D. I. V., and A. H. Butler, 2020: Stratospheric drivers of extreme events at the Earth's surface. *Commun. Earth Environ.*, **1**, 59, https://doi.org/10.1038/s43247-020-00060-z.

Donlon, C. J., M. Martin, J. Stark, J. Roberts-Jones, E. Fiedler, and W. Wimmer, 2012: The Operational Sea Surface Temperature and Sea Ice Analysis (OSTIA) system. *Remote Sens. Environ.*, **116**, 140–158, https://doi.org/10.1016/j.rse.2010.10.017.

Dörnbrack, A., M. Leutbecher, R. Kivi, and E. Kyrö, 1999: Mountain-wave-induced record low stratospheric temperatures above northern Scandinavia. *Tellus*, **51A**, 951–963, https://doi.org/10.3402/tellusa.v51i5.14504.

——, S. Gisinger, M. C. Pitts, L. R. Poole, and M. Maturilli, 2017: Multilevel cloud structures over Svalbard. *Mon. Wea. Rev.*, **145**, 1149–1159, https://doi.org/10.1175/MWR-D-16-0214.1.

Eckermann, S. D., and Coauthors, 2018: High-altitude (0–100 km) global atmospheric reanalysis system: Description and application to the 2014 austral winter of the Deep Propagating Gravity Wave Experiment (DEEPWAVE). *Mon. Wea. Rev.*, **146**, 2639–2666, https://doi.org/10.1175/MWR-D-17-0386.1.

Ehard, B., B. Kaifler, N. Kaifler, and M. Rapp, 2015: Evaluation of methods for gravity wave extraction from middle-atmospheric lidar temperature measurements. *Atmos. Meas. Tech.*, **8**, 4645–4655, https://doi.org/10.5194/amt-8-4645-2015.

Eichinger, R., H. Garny, P. Šácha, J. Danker, S. Dietmüller, and S. Oberländer-Hayn, 2020: Effects of missing gravity waves on stratospheric dynamics; Part 1: Climatology. *Climate Dyn.*, **54**, 3165–3183, https://doi.org/10.1007/s00382-020-05166-w.

——, and Coauthors, 2024: Emulating lateral gravity wave propagation in a global chemistry-climate model (EMAC v2.55.2) through horizontal flux redistribution. *EGUsphere*, https://doi.org/10.5194/egusphere-2023-270, in press.

Eliassen, A., and E. Palm, 1961: On the transfer of energy in stationary mountain waves. *Geofys. Publ.*, **22**, 1–23.

Ern, M., P. Preusse, M. J. Alexander, and C. D. Warner, 2004: Absolute values of gravity wave momentum flux derived from satellite data. *J. Geophys. Res.*, **109**, D20103, https://doi.org/10.1029/2004JD004752.

——, and Coauthors, 2014: Interaction of gravity waves with the QBO: A satellite perspective. *J. Geophys. Res. Atmos.*, **119**, 2329–2355, https://doi.org/10.1002/2013JD020731.

——, 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**, 475–485, https://doi.org/10.1002/2016GL072007.

——, Q. T. Trinh, P. Preusse, J. C. Gille, M. G. Mlynczak, J. M. Russell III, and M. Riese, 2018: GRACILE: A comprehensive climatology of atmospheric gravity wave parameters based on satellite limb soundings. *Earth Syst. Sci. Data*, **10**, 857–892, https://doi.org/10.5194/essd-10-857-2018.

Fritts, D. C., and M. J. Alexander, 2003: Gravity wave dynamics and effects in the middle atmosphere. *Rev. Geophys.*, **41**, 1003, https://doi.org/10.1029/2001RG000106.

Garcia, R. R., and S. Solomon, 1985: The effect of breaking gravity waves on the dynamics and chemical composition of the

mesosphere and lower thermosphere. *J. Geophys. Res.*, **90**, 3850–3868, https://doi.org/10.1029/JD090iD02p03850.

Geller, M. A., and Coauthors, 2013: A comparison between gravity wave momentum fluxes in observations and climate models. *J. Climate*, **26**, 6383–6405, https://doi.org/10.1175/JCLI-D-12-00545.1.

Gill, A. E., 1982: *Atmosphere–Ocean Dynamics.* Academic Press, 662 pp.

Giorgetta, M. A., E. Manzini, and E. Roeckner, 2002: Forcing of the quasi-biennial oscillation from a broad spectrum of atmospheric waves. *Geophys. Res. Lett.*, **29**, 1245, https://doi.org/10.1029/2002GL014756.

Gisinger, S., and Coauthors, 2017: Atmospheric conditions during the Deep Propagating Gravity Wave Experiment (DEEP-WAVE). *Mon. Wea. Rev.*, **145**, 4249–4275, https://doi.org/10.1175/MWR-D-16-0435.1.

——, I. Polichtchouk, A. Dörnbrack, R. Reichert, B. Kaifler, N. Kaifler, M. Rapp, and I. Sandu, 2022: Gravity-wave-driven seasonal variability of temperature differences between ECMWF IFS and Rayleigh lidar measurements in the lee of the southern Andes. *J. Geophys. Res. Atmos.*, **127**, e2021JD036270, https://doi.org/10.1029/2021JD036270.

Granier, C., and Coauthors, 2011: Evolution of anthropogenic and biomass burning emissions of air pollutants at global and regional scales during the 1980–2010 period. *Climatic Change*, **109**, 163, https://doi.org/10.1007/s10584-011-0154-1.

Gupta, A., T. Birner, A. Dörnbrack, and I. Polichtchouk, 2021: Importance of gravity wave forcing for springtime southern polar vortex breakdown as revealed by ERA5. *Geophys. Res. Lett.*, **48**, e2021GL092762, https://doi.org/10.1029/2021GL092762.

Hendricks, E. A., J. D. Doyle, S. D. Eckermann, Q. Jiang, and P. A. Reinecke, 2014: What is the source of the stratospheric gravity wave belt in austral winter? *J. Atmos. Sci.*, **71**, 1583–1592, https://doi.org/10.1175/JAS-D-13-0332.1.

Hersbach, H., and Coauthors, 2020: The ERA5 global reanalysis. *Quart. J. Roy. Meteor. Soc.*, **146**, 1999–2049, https://doi.org/10.1002/qj.3803.

Hertzog, A., G. Boccara, R. A. Vincent, F. Vial, and P. Cocquerez, 2008: Estimation of gravity wave momentum flux and phase speeds from quasi-Lagrangian stratospheric balloon flights. Part II: Results from the Vorcore campaign in Antarctica. *J. Atmos. Sci.*, **65**, 3056–3070, https://doi.org/10.1175/2008JAS2710.1.

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**, e2020GL089557, https://doi.org/10.1029/2020GL089557.

——, and Coauthors, 2021: Stratospheric gravity waves over the mountainous island of South Georgia: Testing a high-resolution dynamical model with 3-D satellite observations and radiosondes. *Atmos. Chem. Phys.*, **21**, 7695–7722, https://doi.org/10.5194/acp-21-7695-2021.

Hines, C. O., 1997: Doppler-spread parameterization of gravity-wave momentum deposition in the middle atmosphere. Part 2: Broad and quasi monochromatic spectra, and implementation. *J. Atmos. Sol.-Terr. Phys.*, **59**, 387–400, https://doi.org/10.1016/S1364-6826(96)00080-6.

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. Atmos.*, **118**, 416–434, https://doi.org/10.1029/2012JD018658.

——, 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**, 2901–2920, https://doi.org/10.5194/acp-17-2901-2017.

Holton, J. R., 1982: The role of gravity wave induced drag and diffusion in the momentum budget of the mesosphere. *J. Atmos. Sci.*, **39**, 791–799, https://doi.org/10.1175/1520-0469(1982)039<0791:TROGWI>2.0.CO;2.

Höpfner, M., and Coauthors, 2006: MIPAS detects Antarctic stratospheric belt of NAT PSCs caused by mountain waves. *Atmos. Chem. Phys.*, **6**, 1221–1230, https://doi.org/10.5194/acp-6-1221-2006.

Hourdin, F., and Coauthors, 2017: The art and science of climate model tuning. *Bull. Amer. Meteor. Soc.*, **98**, 589–602, https://doi.org/10.1175/BAMS-D-15-00135.1.

Jöckel, P., and Coauthors, 2010: Development cycle 2 of the Modular Earth Submodel System (MESSy2). *Geosci. Model Dev.*, **3**, 717–752, https://doi.org/10.5194/gmd-3-717-2010.

——, and Coauthors, 2016: Earth System Chemistry integrated Modelling (ESCiMo) with the Modular Earth Submodel System (MESSy) version 2.51. *Geosci. Model Dev.*, **9**, 1153–1200, https://doi.org/10.5194/gmd-9-1153-2016.

Kaifler, B., and N. Kaifler, 2021: A Compact Rayleigh Autonomous Lidar (CORAL) for the middle atmosphere. *Atmos. Meas. Tech.*, **14**, 1715–1732, https://doi.org/10.5194/amt-14-1715-2021.

——, ——, B. Ehard, A. Dörnbrack, M. Rapp, and D. C. Fritts, 2015: Influences of source conditions on mountain wave penetration into the stratosphere and mesosphere. *Geophys. Res. Lett.*, **42**, 9488–9494, https://doi.org/10.1002/2015GL066465.

Kaifler, N., B. Kaifler, A. Dörnbrack, M. Rapp, J. L. Hormaechea, and A. de la Torre, 2020: Lidar observations of large-amplitude mountain waves in the stratosphere above Tierra del Fuego, Argentina. *Sci. Rep.*, **10**, 14529, https://doi.org/10.1038/s41598-020-71443-7.

Kidston, J., A. A. Scaife, S. C. Hardiman, D. M. Mitchell, N. Butchart, M. P. Baldwin, and L. J. Gray, 2015: Stratospheric influence on tropospheric jet streams, storm tracks and surface weather. *Nat. Geosci.*, **8**, 433–440, https://doi.org/10.1038/ngeo2424.

Kim, Y.-J., S. D. Eckermann, and H.-Y. Chun, 2003: An overview of the past, present and future of gravity-wave drag parametrization for numerical climate and weather prediction models. *Atmos.–Ocean*, **41**, 65–98, https://doi.org/10.3137/ao.410105.

Klaver, R., R. Haarsma, P. L. Vidale, and W. Hazeleger, 2020: Effective resolution in high resolution global atmospheric models for climate studies. *Atmos. Sci. Lett.*, **21**, e952, https://doi.org/10.1002/asl.952.

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

——, and Coauthors, 2022: Observed and modeled mountain waves from the surface to the mesosphere near the Drake Passage. *J. Atmos. Sci.*, **79**, 909–932, https://doi.org/10.1175/JAS-D-21-0252.1.

Lindgren, E. A., A. Sheshadri, A. Podglajen, and R. W. Carver, 2020: Seasonal and latitudinal variability of the gravity wave spectrum in the lower stratosphere. *J. Geophys. Res. Atmos.*, **125**, e2020JD032850, https://doi.org/10.1029/2020JD032850.

Lott, F., and M. J. Miller, 1997: A new subgrid-scale orographic drag parametrization: Its formulation and testing. *Quart. J. Roy. Meteor. Soc.*, **123**, 101–127, https://doi.org/10.1002/qj.4971235 3704.

Maciel, P., T. Quintino, U. Modigliani, P. Dando, B. Raoult, W. Deconinck, F. Rathgeber, and C. Simarro, 2017: The new ECMWF interpolation package MIR. *ECMWF Newsletter*,

No. 152, ECMWF, Reading, United Kingdom, 36–39, https://www.ecmwf.int/en/elibrary/80568-new-ecmwf-interpolation-package-mir.

Mansfield, L. A., and A. Sheshadri, 2022: Calibration and uncertainty quantification of a gravity wave parameterization: A case study of the quasi-biennial oscillation in an intermediate complexity climate model. *J. Adv. Model. Earth Syst.*, **14**, e2022MS003245, https://doi.org/10.1029/2022MS003245.

McLandress, C., J. F. Scinocca, T. G. Shepherd, M. C. Reader, and G. L. Manney, 2013: Dynamical control of the mesosphere by orographic and nonorographic gravity wave drag during the extended northern winters of 2006 and 2009. *J. Atmos. Sci.*, **70**, 2152–2169, https://doi.org/10.1175/JAS-D-12-0297.1.

Minamihara, Y., K. Sato, and M. Tsutsumi, 2020: Intermittency of gravity waves in the Antarctic troposphere and lower stratosphere revealed by the PANSY radar observation. *J. Geophys. Res. Atmos.*, **125**, e2020JD032543, https://doi.org/10.1029/2020JD032543.

Morfa, Y. A., and C. C. Stephan, 2023: The relationship between horizontal and vertical velocity wavenumber spectra in global storm-resolving simulations. *J. Atmos. Sci.*, **80**, 1087–1105, https://doi.org/10.1175/JAS-D-22-0105.1.

Morgenstern, O., and Coauthors, 2017: Review of the global models used within phase 1 of the Chemistry–Climate Model Initiative (CCMI). *Geosci. Model Dev.*, **10**, 639–671, https://doi.org/10.5194/gmd-10-639-2017.

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**, 5905–5926, https://doi.org/10.1175/2010JCLI3490.1.

Pahlavan, H. A., J. M. Wallace, and Q. Fu, 2023: Characteristics of tropical convective gravity waves resolved by ERA5 reanalysis. *J. Atmos. Sci.*, **80**, 777–795, https://doi.org/10.1175/JAS-D-22-0057.1.

Plougonven, R., A. Hertzog, and L. Guez, 2013: Gravity waves over Antarctica and the Southern Ocean: Consistent momentum fluxes in mesoscale simulations and stratospheric balloon observations. *Quart. J. Roy. Meteor. Soc.*, **139**, 101–118, https://doi.org/10.1002/qj.1965.

——, A. de la Cámara, A. Hertzog, and F. Lott, 2020: How does knowledge of atmospheric gravity waves guide their parameterizations? *Quart. J. Roy. Meteor. Soc.*, **146**, 1529–1543, https://doi.org/10.1002/qj.3732.

Polichtchouk, I., T. G. Shepherd, R. J. Hogan, and P. Bechtold, 2018: Sensitivity of the Brewer–Dobson circulation and polar vortex variability to parameterized nonorographic gravity wave drag in a high-resolution atmospheric model. *J. Atmos. Sci.*, **75**, 1525–1543, https://doi.org/10.1175/JAS-D-17-0304.1.

——, N. Wedi, and Y.-H. Kim, 2022: Resolved gravity waves in the tropical stratosphere: Impact of horizontal resolution and deep convection parametrization. *Quart. J. Roy. Meteor. Soc.*, **148**, 233–251, https://doi.org/10.1002/qj.4202.

——, A. van Niekerk, and N. Wedi, 2023: Resolved gravity waves in the extratropical stratosphere: Effect of horizontal resolution increase from $O(10)$ to $O(1)$ km. *J. Atmos. Sci.*, **80**, 473–486, https://doi.org/10.1175/JAS-D-22-0138.1.

Preusse, P., M. Ern, P. Bechtold, S. D. Eckermann, S. Kalisch, Q. T. Trinh, and M. Riese, 2014: Characteristics of gravity waves resolved by ECMWF. *Atmos. Chem. Phys.*, **14**, 10 483–10 508, https://doi.org/10.5194/acp-14-10483-2014.

Procházková, Z., and Coauthors, 2023: Sensitivity of mountain wave drag estimates on separation methods and proposed improvements. *J. Atmos. Sci.*, **80**, 1661–1680, https://doi.org/10.1175/JAS-D-22-0151.1.

Queney, P., 1948: The problem of air flow over mountains: A summary of theoretical studies. *Bull. Amer. Meteor. Soc.*, **29**, 16–26, https://doi.org/10.1175/1520-0477-29.1.16.

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

Rapp, M., A. Dörnbrack, and P. Preusse, 2018: Large midlatitude stratospheric temperature variability caused by inertial instability: A potential source of bias for gravity wave climatologies. *Geophys. Res. Lett.*, **45**, 10 682–10 690, https://doi.org/10.1029/2018GL079142.

——, 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**, E871–E893, https://doi.org/10.1175/BAMS-D-20-0034.1.

Rayner, N. A., D. E. Parker, E. B. Horton, C. K. Folland, L. V. Alexander, D. P. Rowell, E. C. Kent, and A. Kaplan, 2003: Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *J. Geophys. Res.*, **108**, 4407, https://doi.org/10.1029/2002JD002670.

Reichert, R., B. Kaifler, N. Kaifler, A. Dörnbrack, M. Rapp, and J. L. Hormaechea, 2021: High-cadence lidar observations of middle atmospheric temperature and gravity waves at the southern Andes hot spot. *J. Geophys. Res. Atmos.*, **126**, e2021JD034683, https://doi.org/10.1029/2021JD034683.

Roeckner, E., and Coauthors, 2003: The atmospheric general circulation model ECHAM 5. Part I: Model description. Max-Planck-Institut für Meteorologie Rep. 349, 127 pp., https://doi.org/10.17617/2.995269.

Sato, K., and S. Hirano, 2019: The climatology of the Brewer–Dobson circulation and the contribution of gravity waves. *Atmos. Chem. Phys.*, **19**, 4517–4539, https://doi.org/10.5194/acp-19-4517-2019.

——, S. Watanabe, Y. Kawatani, Y. Tomikawa, K. Miyazaki, and M. Takahashi, 2009: On the origins of mesospheric gravity waves. *Geophys. Res. Lett.*, **36**, L19801, https://doi.org/10.1029/2009GL039908.

——, 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**, 1378–1396, https://doi.org/10.1175/JAS-D-11-0101.1.

Schroeder, S., P. Preusse, M. Ern, and M. Riese, 2009: Gravity waves resolved in ECMWF and measured by SABER. *Geophys. Res. Lett.*, **36**, L10805, https://doi.org/10.1029/2008GL037054.

Schulzweida, U., 2005: CDO user's guide. MPI, https://code.mpimet.mpg.de/projects/cdo/embedded/index.html.

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

Shutts, G. J., and S. B. Vosper, 2011: Stratospheric gravity waves revealed in NWP model forecasts. *Quart. J. Roy. Meteor. Soc.*, **137**, 303–317, https://doi.org/10.1002/qj.763.

Skamarock, W. C., 2004: Evaluating mesoscale NWP models using kinetic energy spectra. *Mon. Wea. Rev.*, **132**, 3019–3032, https://doi.org/10.1175/MWR2830.1.

Song, B.-G., H.-Y. Chun, and I.-S. Song, 2020: Role of gravity waves in a vortex-split sudden stratospheric warming in

January 2009. *J. Atmos. Sci.*, **77**, 3321–3342, https://doi.org/10.1175/JAS-D-20-0039.1.

Stephan, C. C., C. Strube, D. Klocke, M. Ern, L. Hoffmann, P. Preusse, and H. Schmidt, 2019: Intercomparison of gravity waves in global convection-permitting models. *J. Atmos. Sci.*, **76**, 2739–2759, https://doi.org/10.1175/JAS-D-19-0040.1.

——, and Coauthors, 2022: Atmospheric energy spectra in global kilometre-scale models. *Tellus*, **74A**, 280–299, https://doi.org/10.16993/tellusa.26.

Strube, C., M. Ern, P. Preusse, and M. Riese, 2020: Removing spurious inertial instability signals from gravity wave temperature perturbations using spectral filtering methods. *Atmos. Meas. Tech.*, **13**, 4927–4945, https://doi.org/10.5194/amt-13-4927-2020.

Torrence, C., and G. P. Compo, 1998: A practical guide to wavelet analysis. *Bull. Amer. Meteor. Soc.*, **79**, 61–78, https://doi.org/10.1175/1520-0477(1998)079<0061:APGTWA>2.0.CO;2.

van Niekerk, A., and S. B. Vosper, 2021: Towards a more "scale-aware" orographic gravity wave drag parametrization: Description and initial testing. *Quart. J. Roy. Meteor. Soc.*, **147**, 3243–3262, https://doi.org/10.1002/qj.4126.

——, and Coauthors, 2020: Constraining Orographic Drag Effects (COORDE): A model comparison of resolved and parametrized orographic drag. *J. Adv. Model. Earth Syst.*, **12**, e2020MS002160, https://doi.org/10.1029/2020MS002160.

Voelker, G. S., G. Bölöni, Y.-H. Kim, G. Zängl, and U. Achatz, 2023: MS-GWaM: A 3-dimensional transient gravity wave parametrization for atmospheric models. arXiv, 2309.11257v1, https://doi.org/10.48550/arXiv.2309.11257.

Wedi, N. P., and Coauthors, 2020: A baseline for global weather and climate simulations at 1 km resolution. *J. Adv. Model. Earth Syst.*, **12**, e2020MS002192, https://doi.org/10.1029/2020MS002192.

Weimer, M., C. Wilka, D. E. Kinnison, R. R. Garcia, J. T. Bacmeister, M. J. Alexander, A. Dörnbrack, and S. Solomon, 2023: A method for estimating global subgrid-scale orographic gravity-wave temperature perturbations in chemistry-climate models. *J. Adv. Model. Earth Syst.*, **15**, e2022MS003505, https://doi.org/10.1029/2022MS003505.

Wilson, R., M. L. Chanin, and A. Hauchecorne, 1991: Gravity waves in the middle atmosphere observed by Rayleigh lidar: 2. Climatology. *J. Geophys. Res.*, **96**, 5169–5183, https://doi.org/10.1029/90JD02610.

Xu, J., A. K. Smith, and G. P. Brasseur, 2000: The effects of gravity waves on distributions of chemically active constituents in the mesopause region. *J. Geophys. Res.*, **105**, 26 593–26 602, https://doi.org/10.1029/2000JD900446.

Žagar, N., D. Jelić, M. J. Alexander, and E. Manzini, 2018: Estimating subseasonal variability and trends in global atmosphere using reanalysis data. *Geophys. Res. Lett.*, **45**, 12 999–13 007, https://doi.org/10.1029/2018GL080051.