Gravity-Wave-Driven Seasonal Variability of Temperature Differences between ECMWF IFS and Rayleigh Lidar Measurements in the Lee of the Southern Andes

Sonja Gisinger[1], Inna Polichtchouk[2], Andreas Dörnbrack[1], Robert Reichert[1], Bernd Kaifler[1], Natalie Kaifler[1], Markus Rapp [1,3], Irina Sandu[2]

[1]Deutsches Zentrum für Luft- und Raumfahrt, Institut für Physik der Atmosphäre, Oberpfaffenhofen, Germany
[2]European Centre for Medium-Range Weather Forecasts, Reading, UK
[3]Ludwig-Maximilians-Universität München, Munich, Germany

Key Points:

- Monthly mean temperature differences between the IFS and ground-based lidar measurements in the middle atmosphere are largest in winter.
- Wintertime temperature deviations are related to stratospheric gravity wave activity and both amplitude and phase deviations are important.
- Damping by the model sponge and unresolved gravity waves reduce gravity wave potential energy in the middle atmosphere in the IFS.

Corresponding author: Sonja Gisinger, `sonja.gisinger@dlr.de`

Abstract

Long-term high-resolution data of the Compact Rayleigh Autonomous Lidar (CORAL) is used to evaluate ECMWF Integrated Forecasting Systems' (IFS) temperature data and gravity wave (GW) activity over Rio Grande (53.79°S, 67.75°W) which is a hot spot of stratospheric GWs in winter. Seasonal and altitudinal variations of the temperature differences are studied for 2018 with a uniform IFS version. Moreover, interannual variations are considered using updated versions of the IFS. We find monthly mean temperature differences of less than 2 K and high Pearson correlation coefficients (> 0.7) at 20-40 km. At 45-55 km, the differences are smaller than 4 K during summer conditions. The largest differences are found during winter conditions (4 K in May 2018 and -10 K in August 2018, July 2019 and 2020). The standard deviation of the differences and maximum differences between instantaneous individual profiles are also larger during winter conditions (> ±10 K) and increase with altitude. We relate this seasonal variability of the temperature deviations to stratospheric GW activity. In the upper stratosphere and lower mesosphere, the observed temperature differences result from both GW amplitude and phase deviations. The IFS captures the seasonal cycle of GW potential energy (E_p) well, but underestimates E_p in the middle atmosphere. Experimental IFS runs without damping by the model sponge for May and August 2018 show an increase of the monthly mean E_p above 45 km from only ≈ 10 % of the E_p derived from the lidar measurements to 25 % and 42 %, respectively. GWs not resolved in the IFS are likely explaining the remaining underestimation of the E_p.

1 Introduction

Even nowadays when we have a growing understanding of stratospheric processes, highly developed numerical models, and increasing computational resources, stratospheric temperatures in atmospheric analyses and reanalyses still have a larger uncertainty than their tropospheric counterparts. Improving the representation of the past (reanalysis), current (analysis), and future (forecast) state of the middle atmosphere in general circulation models (GCMs) is important for the validation and forecasting of tropospheric weather and future climate. It is known that the circulation in the middle and upper atmosphere is strongly influenced by internal gravity waves (GWs) triggered for example by flow over mountains (Fritts & Alexander, 2003). At the same time, processes in the stratosphere such as anomalies in the winter- and spring-time stratospheric polar vortex impact the tropospheric circulation (Baldwin & Dunkerton, 2001; Garfinkel & Hartmann, 2011; Byrne & Shepherd, 2018).

One issue when modelling the middle atmosphere is that there is a limited amount of observations to constrain the current state in the model (e.g., Eckermann et al., 2018). Above 10 hPa, most of the observations assimilated into the Integrated Forecasting System (IFS) of the European Centre for Medium-Range Weather Forecasts (ECMWF) are from satellites and have limited spatial and temporal resolutions. They mainly provide temperature-related data (e.g., Global Navigation Satellite System Radio Occultation (GNSS-RO), Atmospheric Infrared Sounder (AIRS), Advanced Microwave Sounding Unit (AMSU-A)) and the topmost radiances assimilated peak at 2 hPa. The range of sensitivity of the satellite observations to certain horizontal and vertical scales of GWs depends on the instrument and viewing geometry (observational filter Alexander, 1998) as can be seen in e.g., Figure 9 of Preusse et al. (2008). To produce the most accurate representation of the atmospheric state, all the observations irregularly distributed in time and space and each having their limitations and uncertainties are combined with the numerical weather prediction model on a global regular grid. For the reanalysis and analysis at ECMWF, this is achieved by 4-dimensional variational data assimilation (4D-Var).

The reanalysis are based on a conserved version of the IFS to reproduce the best possible state of the atmosphere over a past period (e.g. 1950 to present in the case of the fifth's generation of reanalysis (ERA5)). In May 2020, ECMWF released reruns of ERA5 for the years 2000 to 2006 (ERA5.1) in order to improve the global-mean cold bias in the lower stratosphere that was found in ERA5 when compared to observations and previous reanalyses (ERA-Interim) (Simmons et al., 2020). This bias was related to inappropriate background error covariances for the data assimilation for this time period. Another factor contributing to differences between ERA5 and ERA-Interim is the larger lower stratospheric cold model bias in the IFS cycle 41r2, on which ERA5 is based on, compared to cycle 31r2, on which ERA-Interim was based on. The cold model bias switches to a warm bias at higher altitudes in the stratosphere (Simmons et al., 2020) that is related to radiative processes in the model (Shepherd et al., 2018).

The analysis is the best guess of the current atmospheric state that is used to initialize forecasts. Many satellite observations in the upper stratosphere are rejected by the 4D-Var in the IFS over the GW hot spot region of the Southern Andes, the Drake Passage, and the Antarctic Peninsula in the southern hemispheric extended winter period (April to September), most frequently in May (Tony McNally, personal communication, December 2018). The observations deviate too strongly from the IFS background which is likely due to GW-induced temperature perturbations. Stratospheric GW activity doesn't show a homogeneous distribution over the globe but numerous hot spots exist mostly found close to prominent orographic features like mountain ranges, coasts, lakes, deserts, or isolated islands (Hoffmann et al., 2013). For the aforementioned region, backward ray traycing of GWs at 25 km altitude, which are resolved in the IFS in simulated satellite observations imitating an infrared limb imager, revealed the Antarctic Peninsula and the southern part of the Andes as prominent GW sources (Preusse et al., 2014). Together with GWs generated by storms, these GWs are responsible for huge day-to-day variations in the stratospheric GW momentum flux in the southern hemisphere (Preusse et al., 2014).

The sparsity and limitations of observations in the middle atmosphere means that the model plays a larger role in the determination of the atmospheric state in the analysis and reanalysis. To represent stratospheric processes, the model top and corresponding sponge layers have to be moved further and further up to higher altitudes (Shepherd et al., 1996). This and the enhancement of vertical resolution led to an increasing demand of computational resources that only became available in the past decades. For example, the IFS had 31 vertical levels extending from the troposphere to the mid-stratosphere at 10 hPa back in 2003 (ECMWF, 2003). In 2013, the number of vertical levels was increased from 91 to 137 which is still in use today and the model top is in the mesosphere at 0.01 hPa ($\approx$ 80 km). The sponge layer starts weak at 10 hPa ($\approx$ 28 km) and is strongest above 1 hPa ($\approx$ 45 km) where spectral diffusion is applied on vorticity, divergence, and temperature (Ehard et al., 2018). GWs are significantly damped in the sponge layer to reduce reflection at the model top and to enable the linear calculations in the 4D-Var assimilation scheme of the IFS. The sponge layer leads to a misrepresentation of the GW drag which affects the temperature and circulation of the middle atmosphere (Shepherd et al., 1996). Therefore, reducing the depth and the strength of the sponge could help to improve the representation of GWs and temperature biases in the middle atmosphere.

Temperature biases and the challenges of modelling the middle atmosphere that include the representation of physical and dynamical processes, data assimilation, and artificial damping by sponges motivates our study. Local lidar measurements of the middle atmosphere can be used to evaluate IFS based reanalyses, analyses, and forecasts at altitudes where the amount of assimilated data is smallest, the influence of the model sponge is largest, and the model's vertical resolution is reduced. Marlton et al. (2021) compared ERA-Interim and ERA5 stratospheric temperatures with temperature measurements from ground-based lidar at four sites in the northern hemisphere winter for multiple years (1990-2017) and found mean temperature differences of ±5 K. ERA5 temperatures are found to be too low at 1 hPa at all four sites but a different behaviour is found at each site below 1 hPa. Studies that compared lidar measurements in the northern hemisphere with IFS data were published for wind (Khaykin et al., 2020; Rüfenacht et al., 2018) and for temperature data (Le Pichon et al., 2015; Ehard et al., 2018). Le Pichon et al. (2015) compared nightly mean lidar data of wind and temperature at mid-latitudes in Europe with the 0 UTC analyses of the IFS (cycles 38r1 and 38r2; operational June 2012 to June 2013 and June 2013 to November 2013, respectively) for northern winter 2012/2013 and summer 2013. They found the largest deviations and the highest variability of the deviations in winter. At that time of the year, the variability from large-scale planetary waves dominated and a sudden stratospheric warming accompanied by enhanced GW activity took place in January 2013. Above 45 km altitude, the mean temperature difference exceeded -5 K and the 95 % percentile of the distribution was around -30 K (Le Pichon et al., 2015). Ehard et al. (2018) presented estimates of middle atmospheric temperature differences over northernmost Europe for December 2015 of around -8 K for IFS cycle 41r1 (operational May 2015 to March 2016) and up to -20 K for IFS cycle 41r2 (operational March 2016 to November 2016) at altitudes above 40 km. For the Southern Island of New Zealand located at mid-latitudes in the southern hemisphere, wintertime-averaged temperature differences (July to September 2014) between lidar and IFS data (cycle 40r1; operational November 2013 to May 2015) were between -3 and 2 K between 45-60 km altitude and exceeded -10 K at 70 km (Appendix B in Gisinger et al., 2017).

These results exemplify that deviations of model temperatures in the middle atmosphere depend on the season and the location, and can be different compared to global-mean bias characteristics. However, the total of all local deviations determines the global-mean bias characteristics and understanding and quantifying local deviations can help to improve global mean biases. For the stratospheric GW hot spot region of the Southern Andes, a detailed quantification of local deviations between middle atmosphere temperature measurements and IFS temperatures, their vertical structure, and their seasonal and interannual variability is still missing. Further, the contribution of shortcomings in the representation of middle atmosphere GWs in the IFS to site-specific temperature deviations can be studied for this region because GWs are dominating the atmospheric state for several months of the year (Hoffmann et al., 2013). In November 2017, the DLR Institute of Atmospheric Physics deployed the ground-based Compact Rayleigh Autonomous Lidar (CORAL) at Rio Grande at the southern tip of South America in Argentina (B. Kaifler & Kaifler, 2021). The nightly lidar temperature measurements have high temporal (15 min) and vertical (900 m) resolutions between 15-95 km altitude. Comprehensive analyses of the whole three-year data set including GW characteristics are presented by Reichert et al. (2021).

GW activity can be estimated from lidar temperature measurements by means of GW potential energy that is calculated from temperature perturbations relative to the background temperature. GW potential energy is proportional to the GW momentum flux based on linear theory (Ern et al., 2004). Ehard et al. (2018) found that the IFS is capable of reproducing the overall temporal evolution of the GW activity in the stratosphere between 30 and 40 km altitude over northernmost Europe for a four-months-period, but that GW amplitudes are effectively damped by the sponge at higher levels. Also in reanalysis data (Modern-Era Retrospective analysis for Research and Applications (MERRA), ERA5), GW potential energy is found to be lower in the middle atmosphere compared to multi-year lidar measurements from two European stations at higher mid-latitudes and polar latitudes (Strelnikova et al., 2021). For the southern hemisphere, a simplified comparison of GW potential energy from the IFS and from lidar measurements (no one-to-one comparison but different years of IFS and observational data) at two locations in Antarctica (Rothera and South Pole) is presented in Yamashita et al. (2010). The IFS generally captured site-specific seasonal variations of GW potential energy in the stratosphere, that are determined by a winter maximum and summer minimum at Rothera or continuously low values of GW potential energy at the South Pole (Yamashita et al., 2010). Comparisons of three-day averaged GW temperature amplitudes of SABER (Sounding of the Atmosphere Using Broadband Emission Radiometry) and IFS at 30 km showed that the annual cycle and shorter-term variations dominated by mountain waves are well represented in the IFS also for South America, but that temperature amplitudes are underestimated in the IFS (Schroeder et al., 2009). Back then prior to the year 2010, the IFS had 91 vertical layers and a horizontal resolution of approximately 25 km.

In this study, we present a systematic comparison of middle atmosphere temperatures and GW potential energy of the independent (i.e., not assimilated in the IFS) and high-resolution CORAL lidar data set and operational and experimental IFS simulations for Rio Grande, a site at higher mid-latitudes (53.79°S, 67.75°W) in the southern hemisphere in the lee of the Southern Andes and a hot spot of stratospheric GWs in winter (Hoffmann et al., 2013). Temperature deviations between lidar and the IFS and seasonal variability of the deviations are investigated in detail. This includes the role of wintertime GW representation and phase deviations in the middle atmosphere in the IFS which are only accessible due to the high temporal resolution of the lidar data allowing a one-to-one comparison of quasi-instantaneous values. The annual cycle of GW activity in the middle and upper stratosphere over Rio Grande in the IFS is compared to the one derived from the lidar observations. The results for temperature deviations and GW activity are then combined to investigate the hypothesis that the seasonal variability of the temperature deviations over Rio Grande is related to the stratospheric GW activity. For two selected months with enhanced GW activity (May and August 2018), the importance of individual strong GW events for the mean monthly GW potential energy in the middle atmosphere in the observations and the IFS is evaluated (i.e. intermittency of GW activity). Finally, the effect of damping by the sponge on GW potential energy in the middle atmosphere over Rio Grande is quantified in experimental IFS simulations without a sponge layer for these two months.

Section 2 describes the lidar system CORAL, its temperature data taken at Rio Grande, the IFS model data, and how the data is analyzed and compared. Results are presented in section 3 and discussed and summarized in section 4.

2 Data and methods

2.1 Lidar system and data

CORAL (B. Kaifler & Kaifler, 2021) uses a 12-W-laser beam at 532 nm wavelength and a 1-m-diameter telescope installed in an 8 ft container for night-time, autonomous atmospheric soundings. Backscattered photons are detected with three height-cascaded elastic detector channels and one Raman channel. Density and temperature profiles on a 100 m vertical grid between 15 and 95 km altitude are determined by top-down integration of the hydrostatic equation every 5 minutes using an integration window of 15 minutes and 900 m vertical smoothing for an adequate signal to noise ratio. The precision for temperature is better than 1 K between 35 and 60 km altitude and typically better than 4 K below 30 km and above 65 km altitude. At times, the measurements at the lowest altitudes are affected by the presence of aerosols. If the aerosol load is too high, corrections are no longer applicable and temperature is underestimated. Such data are omitted by the retrieval algorithm (most frequently < 20 km altitude). To allow for proper statistics at all altitudes for all months, we limit the lowest altitude to 20 km for our analysis.

Measurements with CORAL started at Rio Grande in November 2017. Rio Grande is located in the lee of the Southern Andes at the east coast of Argentina at 100 to 200 km distance from the mountains that are to the south and west and at greater distance north-west of Rio Grande (Reichert et al., 2021). The analyses in this study take into account data of the year 2018 which is continuously covered by the lidar measurements and by a uniform version of the IFS (see Sec. 2.2). In addition, data for May and July 2019 and 2020 are also analyzed to investigate interannual variability using updated IFS model versions. We want to point out here, that CORAL measurements are taken fully autonomously with the help of IFS cloud forecasts and a cloud monitoring all-sky camera relying on star detection. Measurements are only possible during cloud-free/patchy conditions and during night time. These are the conditions our results are valid for. Nighttime hours are between 2 and 7 UTC in mid-summer (December) and between 21 and 12 UTC in mid-winter (July). Figure 1a shows the time series of the middle atmosphere temperature measurements from CORAL from 2018 to 2020 as nightly mean values, which are the averaged temperatures over all measurements available each night. The band of highest middle atmosphere temperatures at the stratopause is perturbed by atmospheric waves in the extended winter period (April to September) and at the same time minimum temperatures in the mid-stratosphere are smaller than 200 K (Fig. 1a).

2.2 IFS model and data

IFS cycle 45r1 was already running in its pre-operational phase during the first months of 2018 and eventually became operational in June 2018. Therefore, seasonal variations of the temperature differences between the lidar measurements and the IFS can be investigated based on a uniform version of the IFS for 2018. The updated cycles 46r1 and 47r1 became operational in June 2019 and June 2020, respectively. All three cycles have a horizontal resolution of about 9 km on the cubic octahedral grid (TCo1279). The model top is located at 0.01 hPa ($\sim$80 km). For the 137 vertical layers, the layer thickness gradually increases from $\sim$300 m at $\sim$10 km altitude to $\sim$400 m at $\sim$20 km altitude, and $\sim$2 km at $\sim$60 km altitude. We only use data up to 70 km altitude due to sparse coverage with only three more

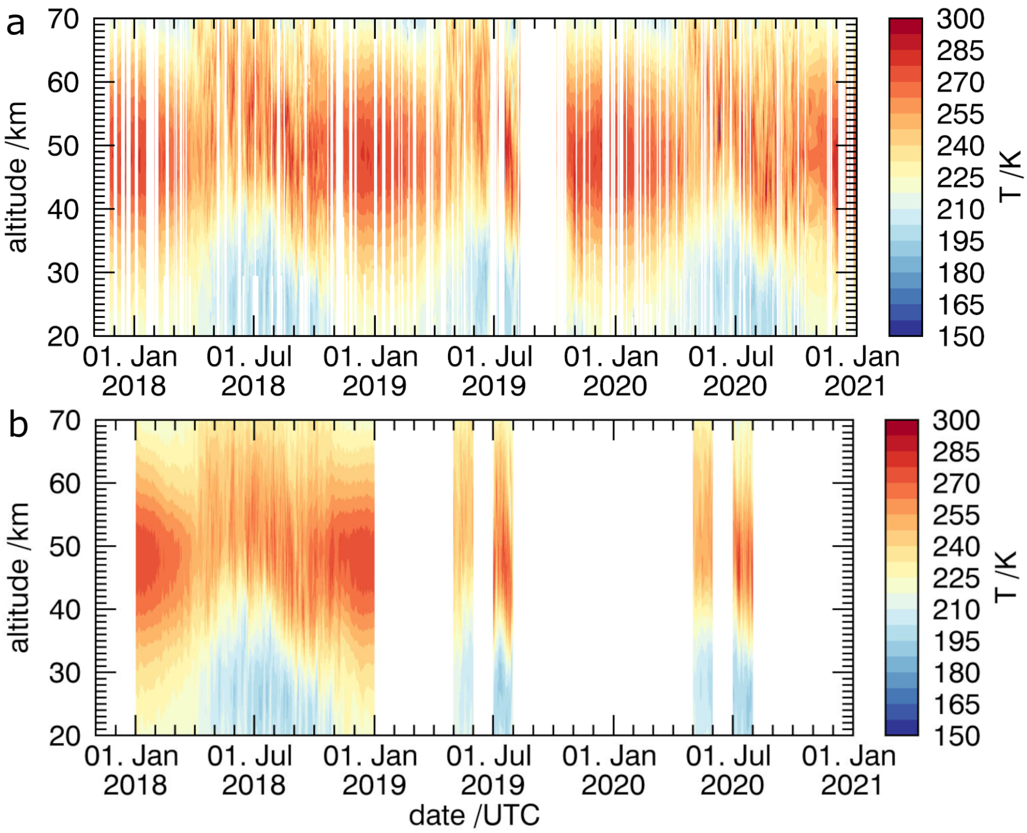

Figure 1. Nightly mean temperatures from CORAL (a) and IFS cycle 45r1 until 11. Jun 2019, cycle 46r1 for July 2019 and May 2020, and cycle 47r1 afterwards (b). Measurement gaps of less than four nights are linearly interpolated in the upper contour plot.

levels above that altitude. In the sponge layer, vertically propagating waves and the zonal mean flow are damped above 10 hPa by hyper-diffusion applied on vorticity, divergence, and temperature and by additional strong first-order damping applied on divergence above 1 hPa. Timescales of both damping mechanisms decrease with altitude and result in stronger damping at the higher altitudes (Polichtchouk et al., 2017; Ehard et al., 2018).

Cycle 46r1 came with changes in the data assimilation (e.g., 4D-Var with an additional outer loop), in the use of observations (e.g., usage of satellites' soil moisture data product and updated geostationary radiances), and in modelling (e.g., new 3D aerosol climatology). On 9 January 2020, ECMWF started assimilating wind data from the Aeolus satellite (Rennie & Isaksen, 2020) in the troposphere and lower stratosphere. For cycle 47r1, the time step in the last minimisation cycle in the 4D-Var is made the same for the outer and inner loop to avoid spurious GW-like increments generated during the 4D-Var analysis. A more detailed description of the changes in the IFS can be found on the ECMWF website (www.ecmwf.int/en/forecasts/documentation-and-support/changes-ecmwf-model, last access Nov 2021). It's not the aim of this paper to quantify changes in middle atmospheric temperatures in the IFS that come with changes in the model setup because this would require the comparison of the different cycles for the same period of time as it was done in Ehard et al. (2018).

IFS analyses are available for 0, 6, 12, and 18 UTC and gaps are filled with short-lead-time forecasts (+1, +2, ...,+5, +7, +8, ..., +11 h) to get hourly data coverage. In addition, experimental forecast runs without the sponge using cycle 45r1 are performed for May and August 2018. These forecasts were run for 48 h and can be directly compared to the operational forecasts with the sponge (up to +11 h). Further, we briefly investigate the effect of longer lead times (+25, ..., +35 h) on the temperature deviations. For best temporal synchronisation, we extract single lidar temperature profiles that are closest in time (max. ±10 min) to each IFS temperature profile at full hour interpolated on the location of Rio Grande. The time step of the IFS (7.5 minutes) is close to the integration window of 15 minutes for the lidar profiles which makes this a reasonable one-to-one comparison. This selection results in a range of 17 (summer) to 183 (winter) temperature profiles per months that are analyzed in this paper (Tab. 1).

In summary, all IFS data of 2018 and May 2019 used in our analysis and investigations are based on operational high-resolution forecast (HRES) data of cycle 45r1 and variability due to fundamental changes in the model itself can be excluded. IFS data for July 2019 and May 2020 are based on cycle 46r1 and finally, data for July 2020 are based on cycle 47r1. Figure 1b shows the timeseries of these IFS temperature data as nightly means taking into account hourly data between 21 and 12 UTC.

2.3 Analysis of temperature differences, GW potential energy, and GW intermittency

The first part of the analysis focuses on the quantification of the temperature deviations and their seasonal and altitudinal variability by calculating the temperature differences between individual lidar and IFS temperature profiles

$$T_{\mathrm{diff}}(z,t) = T_{\mathrm{ECMWF}}(z,t) - T_{\mathrm{lidar}}(z,t), \tag{1}$$

Table 1. Number of profiles per months

year	2018											
month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
#total	19	54	90	117	153	183	162	122	102	69	39	40
#total (20 km)	17	40	73	86	113	170	139	43	28	15	31	33
year	2019						2020					
month	May			Jul			May			Jul		
#total	176			89			146			163		
#total (20 km)	157			69			113			150		

where T_{ECMWF} is the model temperature profile bilinearly interpolated to the horizontal location of the lidar at Rio Grande taking into account the four surrounding grid-points and T_{lidar} is the temperature profile of the lidar. All data are interpolated to a 100 m equidistant grid in altitude (z) and are available in time (t) at full hour. Afterwards monthly means and standard deviation can be calculated

$$\overline{T_{\text{diff}}}(z) = \frac{\sum T_{\text{diff}}(z,t)}{\#\text{total}}, \tag{2}$$

$$T_\sigma(z) = \sqrt{\frac{\sum [T_{\text{diff}}(z,t) - \overline{T_{\text{diff}}}(z)]^2}{\#\text{total} - 1}}, \tag{3}$$

where #total is the number of profiles for each month. The number of profiles at the lowest altitudes can be small for single months because not all measurement profiles reach down to 20 km altitude due to the omittance of lidar data at these altitudes in case of high amounts of aerosols (Sec. 2.1). The number of profiles per months and those reaching down to 20 km altitude are summarized in Table 1. The numbers give an estimate of the amount of profiles that determines the monthly means below and above 30 km altitude. The number of profiles is largest in the extended winter period (April to September) when the nights are longest. They are also included in the relevant figures in Section 3. $\overline{T_{\text{diff}}}(z)$ is equivalent to the difference between the monthly mean temperature profiles (i.e. $\overline{T_{\text{ECMWF}}(z)} - \overline{T_{\text{lidar}}(z)}$). $T_\sigma(z)$ represents the variability of the differences between the individual profiles.

Averaged temperature differences for three altitude ranges (25 to 35 km, 35 to 45 km, and 45 to 55 km) are computed

$$\langle T_{\mathrm{diff}} \rangle_{z_1\text{-}z_2}(t) = \frac{\sum_{z_1}^{z_2} T_{\mathrm{diff}}(z,t)}{n_z}, \tag{4}$$

where n_z is the number of data points in each altitude range (z_1 to z_2). The upper altitude range lies within the IFS' strong sponge layer that starts at 1 hPa (≈ 45 km; Sec. 2.2). The three altitude ranges are evaluated for each month by plotting their histograms with a bin size of 1 K.

The next part of the analysis investigates whether amplitude deviations or misrepresentation of wintertime GWs including phase misalignment determine temperature differences between IFS and lidar data. We here do not go into detail by means of comprehensive case studies but limit the analysis to the simple statistical approach of calculating the Pearson correlation coefficient,

$$PC_{z_1\text{-}z_2}(t) = \frac{\sum_{z_1}^{z_2}[(T_{\mathrm{ECMWF}}(z,t) - \langle T_{\mathrm{ECMWF}}(t)\rangle_{z_1\text{-}z_2})(T_{\mathrm{lidar}}(z,t) - \langle T_{\mathrm{lidar}}(t)\rangle_{z_1\text{-}z_2})]}{\sqrt{\sum_{z_1}^{z_2}(T_{\mathrm{ECMWF}}(z,t) - \langle T_{\mathrm{ECMWF}}(t)\rangle_{z_1\text{-}z_2})^2}\sqrt{\sum_{z_1}^{z_2}(T_{\mathrm{lidar}}(z,t) - \langle T_{\mathrm{lidar}}(t)\rangle_{z_1\text{-}z_2})^2}}, \tag{5}$$

for the mid-stratosphere ($z = 20-40$ km) and the upper stratosphere and lower mesosphere ($z = 40 - 65$ km) for individual temperature profiles within the different months. We only use these two altitude ranges that cover 20 km and 25 km in altitude, respectively, in order to capture larger-scale (> 5 km) structures. For the interpretation of the results, it is assumed that a high PC (> 0.7) stands for good agreement between IFS and lidar in terms of the altitude of temperature maxima and minima (dominated by GWs), i.e. good agreement of the phase alignment. This assumption was verified during inspection of individual profiles when the threshold for good correlation was determined. Moreover, we expect that a low PC in the upper altitude range is accompanied by a low PC in the lower altitude range in the case of a general misrepresentation of upward propagating wintertime GWs the middle atmosphere in the IFS.

Last but not least, GW activity in the IFS is quantified and compared to the lidar measurements by means of GW potential energy per unit mass

$$E_p(z,t) = \frac{1}{2}\frac{g^2}{N^2(z,t)}\frac{\langle T'^2(z,t)\rangle_{15\mathrm{km}}}{T_0^2(z,t)} \tag{6}$$

$$\text{with} \quad N^2(z,t) = \frac{g}{T_0(z,t)}\left(\frac{dT_0(z,t)}{dz} + \frac{g}{c_p}\right), \tag{7}$$

where T' is the temperature fluctuation determined by applying a fifth-order Butterworth high-pass filter with a cut-off wavelength of 15 km to individual vertical profiles, T_0 is the remaining background temperature, N is the Brunt-Väisäla frequency, $g = 9.81$ is the acceleration due to gravity, and c_p is the heat capacity of dry air at constant pressure (Ehard et al., 2015, 2018). For a monochromatic wave, E_p is based on T'^2 that is either integrated along height for one wavelength or along time for one wave period (Tsuda et al., 2004). For our individual profiles irregularily distributed in time, we use vertical averaging with a sliding window (Baumgaertner & McDonald, 2007) with a width of 15 km, i.e. the maximum wavelength in the T'-data, which is marked by the angle brackets in Eq. (6). To avoid edge effects, the uppermost and lowermost 5 km of the E_p-profiles are discarded (Ehard et al., 2015). We limit our comparison and discussion to E_p and do not consider the vertical flux of horizontal momentum because the horizontal wavenumber needed in the computation (Ern et al., 2004; N. Kaifler et al., 2020) is not available from ground-based lidar measurements and corresponding vertical IFS profiles.

The annual cycle of E_p is analyzed for an altitude range between 35 and 50 km (chosen to be the same as in Reichert et al. (2021)) that covers parts of the middle and the upper stratosphere. The distributions of E_p values are determined for the altitude ranges 30 to 45 km and 45 to 60 km for May and August 2018. It has been previously found that stratospheric E_p and GW momentum fluxes show a log-normal distribution rather than a normal distribution (Baumgaertner & McDonald, 2007; Hertzog et al., 2012). The probability density function for the log-normal distribution is given by

$$y = \frac{1}{x\sigma\sqrt{2\pi}} e^{-(\ln x - \mu)^2/2\sigma^2}, \tag{8}$$

where μ is the expected value and σ is the geometric standard deviation (Baumgaertner & McDonald, 2007). Taking this into account, monthly mean $\overline{E_p}$ for an altitude range and $\overline{E_p}$-profiles are given based on the logarithmic mean (or geometric mean of the log-normal distribution) of E_p

$$\overline{E_p} = e^{\hat{\mu}}, \tag{9}$$

$$\hat{\mu} = \frac{\sum \ln[E_p(z,t)]}{n} \tag{10}$$

$$\text{and} \quad \hat{\sigma}^2 = \frac{\sum(\ln[E_p(z,t)] - \hat{\mu})^2}{n} \tag{11}$$

(Baumgaertner & McDonald, 2007) where here $E_p(z,t)$ represents either all (n) values in an particular altitude range to get the monthly means ($\overline{E_p}$) for a height region or all values at each altitude ($n = \#\text{total}$) to get monthly mean $\overline{E_p}$-profiles.

However, distributions of GW activity over mountainous regions may have even larger tails that are not properly described by a log-normal distribution (Plougonven et al., 2013). This enhanced intermittency of GW activity is caused by more frequent extreme GW events over mountainous regions compared to flat landscapes and ocean surfaces. The intermittency of GWs is important because the vertical profiles of GW momentum flux convergence in the middle atmosphere, that determine the wave forcing to the mean wind, is different for sporadic GWs with large amplitudes compared to GWs with same mean momentum but smaller amplitudes (Minamihara et al., 2020). GW intermittency can be well quantified with the Gini coefficient, popular in economics, as proposed and evaluated by Plougonven et al. (2013) by means of GW momentum flux

$$I_g = \frac{\sum_{n=1}^{N-1}(n\bar{f} - F_n)}{\sum_{n=1}^{N-1} n\bar{f}}, \tag{12}$$

where in our case, F_n is the cumulative sum of $E_p(z,t)$ values sorted in ascending order having an average $\bar{f} = F_N/N$. I_g is 0 for a constant time series and it is 1 for a most intermittent series. Near orography (e.g., the Antarctic Peninsula) enhanced values of 0.6-0.7 were found in the lower stratosphere in mesoscale simulations for austral spring 2005 (Plougonven et al., 2013).

3 Results

3.1 Temperature deviations and seasonal variability

This first part of the results section quantifies the temperature differences between CORAL and IFS analyses and short-lead-time forecasts (Eq. (2) and (3)), their altitudinal structure, and their seasonal variability, i.e. how they compare between the extended summer period (October to March) and the extended winter period (April to September), i.e. the GW-active season. Monthly mean temperature differences for 2018 are overall smaller than 2 K in the middle stratosphere below 40 km altitude (Fig. 2). Although a reduced number of data profiles is available at these altitudes (Sec. 2.1), the results suggest slightly cold biased IFS temperatures below 30 km above Rio Grande from March to September 2018 with the largest deviations in August. While most of these months show this tendency for too cold temperatures up to 45 km, temperatures are overestimated by 2 K at 40 km altitude in August 2018. Around the stratopause between 45 and 55 km altitude, the sign of the temperature deviations is changing throughout the year. The largest positive deviation (4 K) occurs in May 2018, the largest negative deviation (-10 K) in August 2018, and un-

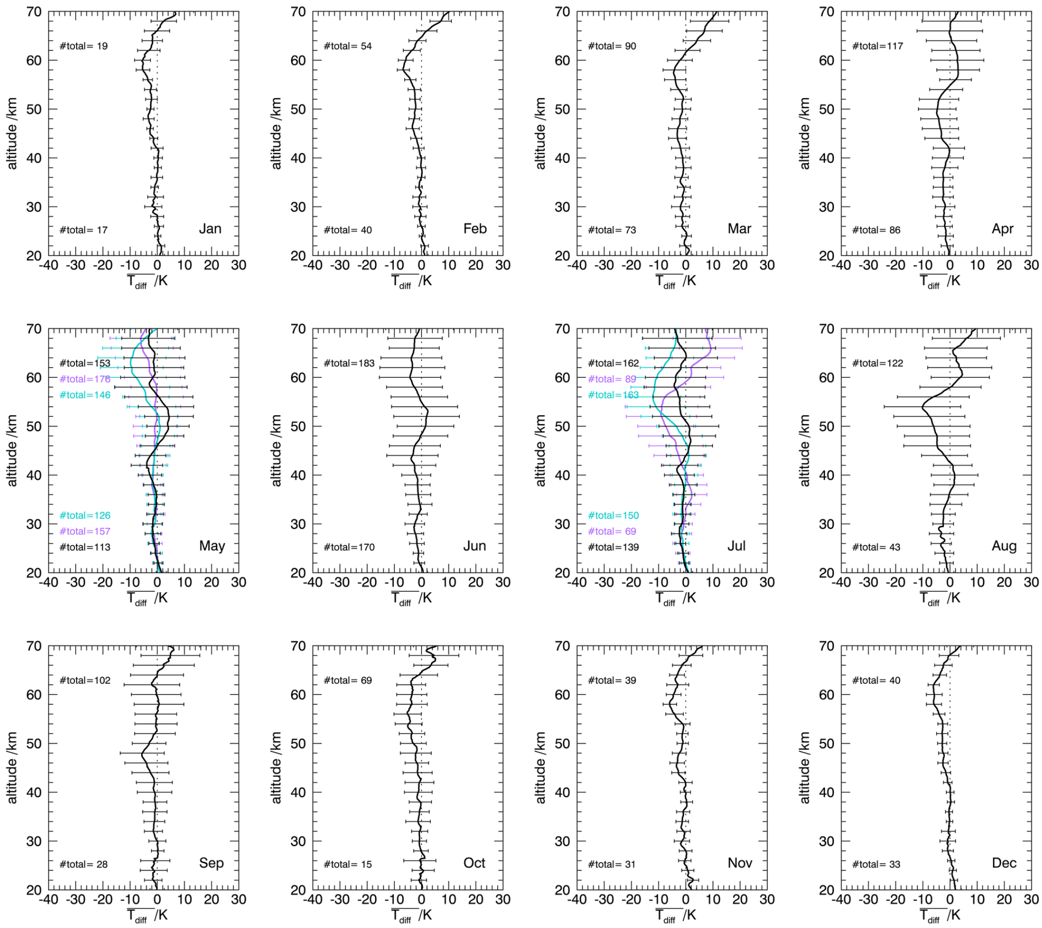

Figure 2. Monthly mean temperature differences and standard deviation between lidar and IFS cycle 45r1 for 2018 (black), for May 2019 (purple), cycle 46r1 for May 2020 (turquois) and July 2019 (purple), and cycle 47r1 for July 2020 (turquois). The number of profiles at 20 km (50 km) altitude is given at the bottom (top) part of the panels and gives an estimate of the amount of profiles that determines the monthly means below and above 30 km altitude (Tab. 1). Negative (positive) values mean that temperatures in the IFS are underestimated (overestimated).

derestimated temperatures (up to -4 K) are found for the months in the extended summer period. At altitudes above 30 km, the results are most reliable because the uncertainty of the lidar measurements is smaller than 1 K between 30 and 60 km altitude (Sec. 2.1). The comparisons for May and August 2018 were repeated for forecast lead times of 25 to 35 hours, and the positive temperature deviations at 50 km altitude for May and at 40 km altitude for August are found to be larger by 1-3 K (not shown). This indicates a warming of the mid-stratosphere for longer lead times. The standard deviation (Eq. (3)), that describes how much the temperature differences of individual temperature profiles of the IFS and the lidar vary within the month, is significantly larger and increases with altitude in the extended winter period (April to September) compared to the other months (Fig. 2). In the upper stratosphere, the standard deviation is up to around 15 K in August 2018. We follow the hypothesis that the presence of GWs in the middle atmosphere can cause large differences for individual temperature profiles during this time of the year due to amplitude and phase errors (analyzed in the following section).

When other years are considered, the mean temperature deviations in the upper stratosphere between 40-50 km become smaller in May 2019 and 2020 in comparison to May 2018 (Fig. 2). It is worth pointing out that the IFS cycle is the same for May 2018 and for May 2019 (Sec. 2) and, hence, the changes cannot be explained by a cycle upgrade of the IFS. For July 2019 and 2020, one finds a large underestimation of IFS temperatures with differences of around -10 K compared to the lidar around the statopause (45 and 50 km altitude), which is not the case for July 2018 but is similar to the deviations found for August 2018 (Fig. 2). The changes could be either due to upgrades of the IFS from cycle 45r1 to 46r1 and 47r1, respectively, or due to variability in the overall atmospheric conditions. Monthly mean stratopause temperatures (not shown) are higher (approx. 268 K) in August 2018, July 2019 and 2020 as compared to July 2018 (approx. 258 K). The IFS does not capture these enhanced stratopause temperatures which explains the larger monthly mean temperature differences between 45 and 55 km altitude for these three months independent of the IFS cycle. Further, the standard deviation in May and July is similar for all three years and increases with altitude also for the updated IFS cycles (46r1, 47r1).

The temperature differences and their variability in the course of the year are investigated in more detail for the three middle-atmospheric altitude ranges (Eq. (4)) by computing histograms. The distribution of the temperature differences is narrowest for the summer months (exemplarily shown for January and October 2018) for all three altitude ranges and differences between individual profiles are rarely found outside the range of ± 5 K (Fig. 3). The largest differences, exceeding ± 5 K, are found in the winter months mainly above 45 km altitude. There, the IFS temperatures can be up to 15 K too warm (May, July 2018) and even more than 15 K too cold (August 2018). The distributions are to a large part identical and similarly broad for May and July 2019 and 2020 (Fig. 4) as in 2018 (Fig. 3). However, the distributions are better centred at zero for May 2019 and 2020 around the stratopause (45-55 km) which results in smaller deviations in the mean profiles shown in Figure 2. In

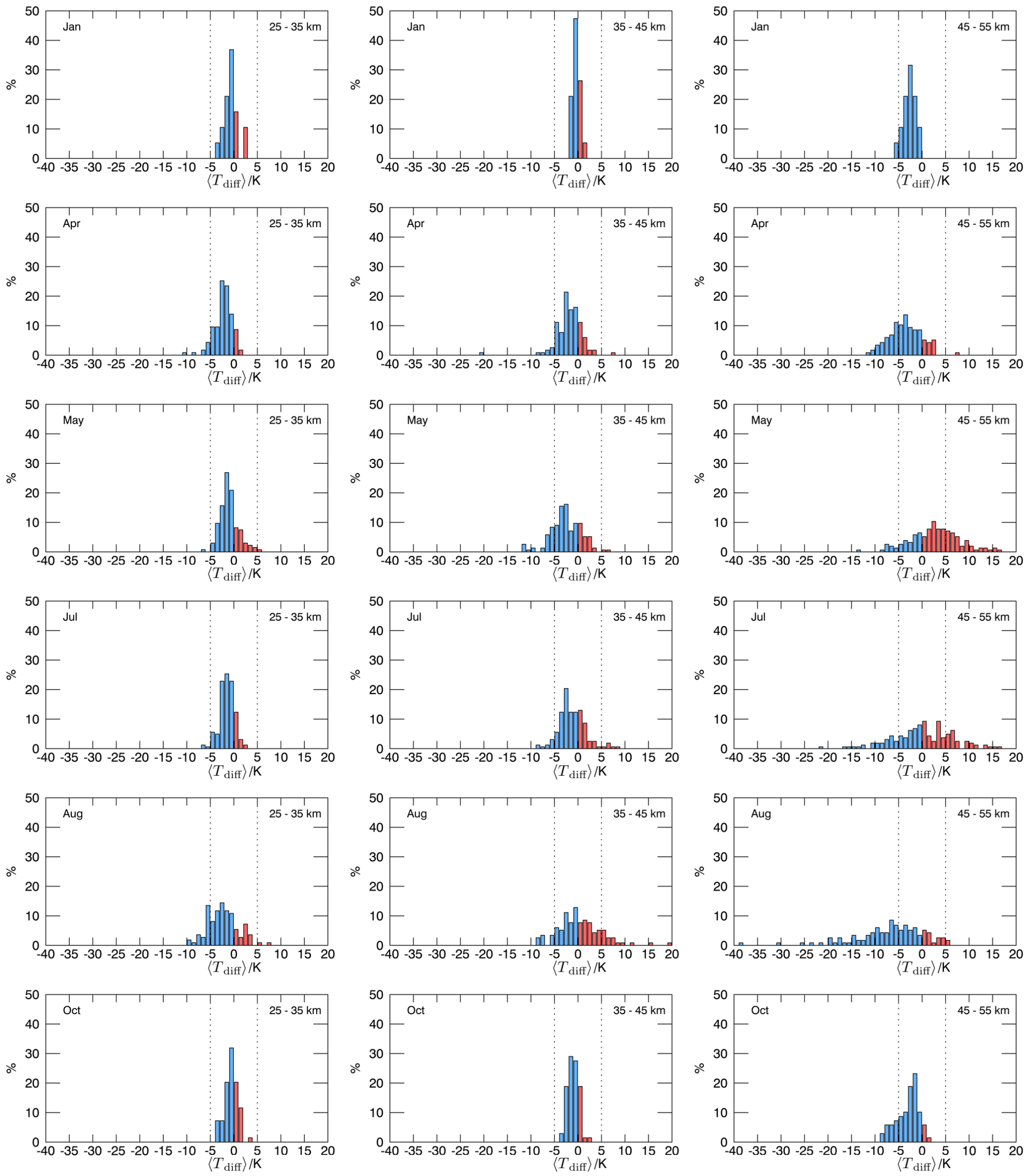

Figure 3. Distribution of temperature differences between lidar and IFS cycle 45r1 for January, April, May, July, August, and October 2018 averaged for 25-35 km altitude (left), 35-45 km altitude (middle), and 45-55 km altitude (right). Negative (positive) temperature differences are blue (red). Vertical dashed lines mark the ± 5 K range.

contrast to that, the distributions for July 2019 and 2020 (Fig. 2) are clearly shifted to negative values in comparison to July 2018 (Fig. 4), i.e. temperatures are more frequently underestimated by more than 5 K as also found for August 2018 (Fig. 3).

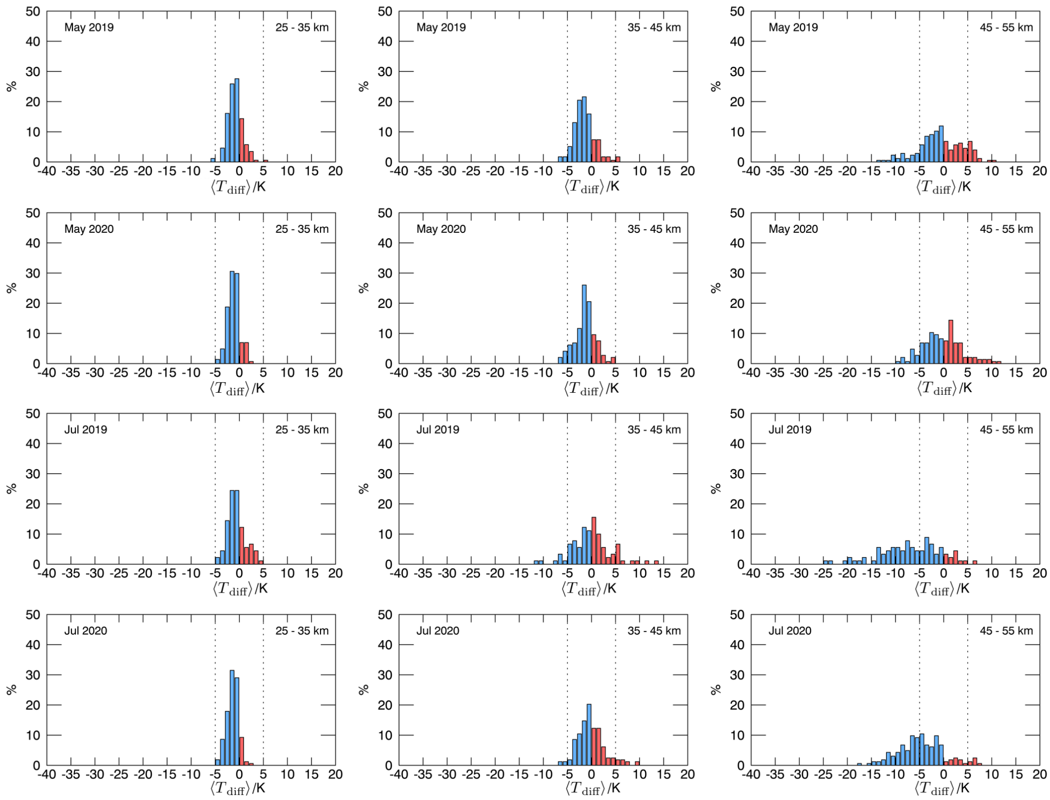

Figure 4. Distribution of temperature differences between lidar and IFS cycle 45r1 for May 2019 (top), cycle 46r1 for May 2020 and July 2019 (middle), and cycle 47r1 for July 2020 (bottom) averaged for 25-35 km altitude (left), 35-45 km altitude (middle), and 45-55 km altitude (right). Negative (positive) temperature differences are blue (red). Vertical dashed lines mark the ± 5 K range.

The annual cycle 2018 of the absolute monthly mean temperature differences ($|\overline{T_{\mathrm{diff}}}|$) and standard deviation (T_σ) averaged for the altitude range of 35 to 50 km is shown in Figure 5. There is no clear annual cycle for the altitude averaged monthly mean temperature differences, especially if May and July 2019 and 2020 are also considered. Minima are found for May 2019 and 2020, and July 2019 because the monthly mean profiles agree well up to 50 km (Fig. 2). However, the standard deviations show maximum values in the extended winter period being continously larger than 4 K which illustrates the seasonal variability found above in the detailed analysis of the individual months. This annual cycle is later correlated to $\overline{E_p}$ of the stratosphere over Rio Grande to relate the seasonal variability of stratospheric temperature deviations to stratospheric GW activity.

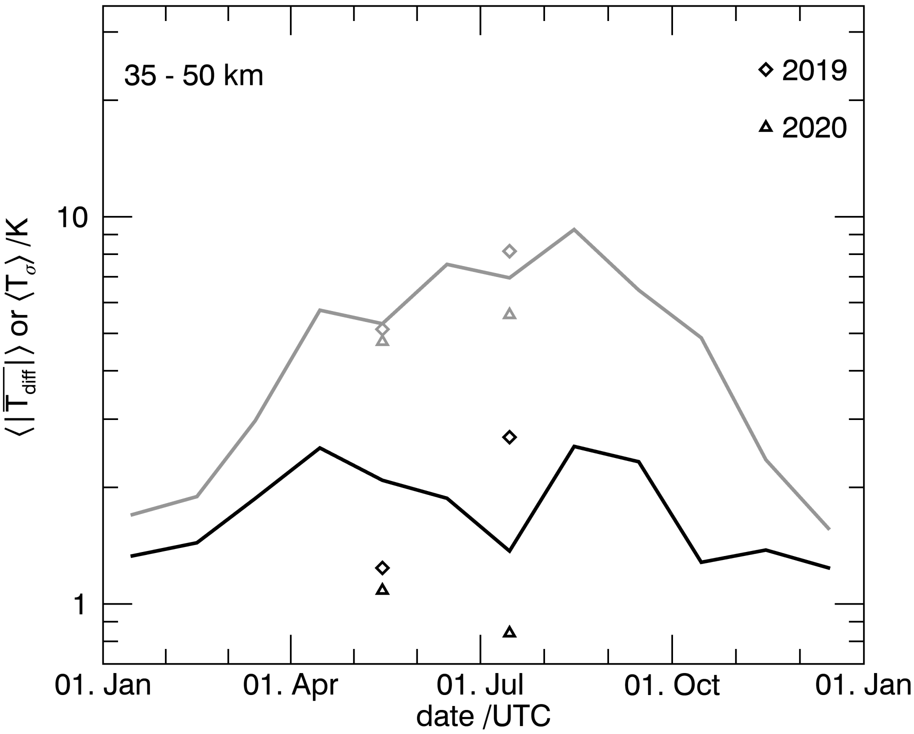

Figure 5. Vertically averaged (35 to 50 km altitude) absolute monthly mean temperature differences (black) between lidar and IFS cycle 45r1 and standard deviation (black) for 2018. Diamonds and triangles are for May and July 2019 and 2020, respectively.

3.2 Amplitude and phase deviations

As the largest temperature differences occur in winter which is the time of enhanced GW activity in the middle atmosphere in the region of Rio Grande (next section and Figure 8), it is now investigated if GW amplitude and/or phase deviations in the IFS are important. Figure 6 shows an example of such amplitude and phase deviations for two individual profiles in May 2018. The profiles for both days show quantitative agreement in phase and amplitude up to 40 km altitude. Higher up, there is an amplitude error of more than 20 K on 31 May 2018 (Fig. 6a) and a phase error of 10 km on 21 May 2018 (Fig. 6b). It was already mentioned that the sponge damps GW amplitudes in the IFS in the stratosphere. Reducing the impact of the sponge may also reduce temperature differences caused by GWs. This is illustrated by the gray profile in Figure 6a where the sponge was removed in the experimental IFS simulations leading to a reduction of the amplitude error at 60 km. However, the removal of the sponge can lead to even larger temperature differences at certain altitudes for cases that show a phase error even though the GW amplitude itself is closer to the observations (gray profile in Figure 6b). In order to quantify the phase agreement, we use PC (Eq. (5)): PC values > 0.7 represent good phase agreement between lidar and IFS (Fig. 6a) and PC values < 0.7 represent poor representation of the GW structure including phase misalignment (Fig. 6b).

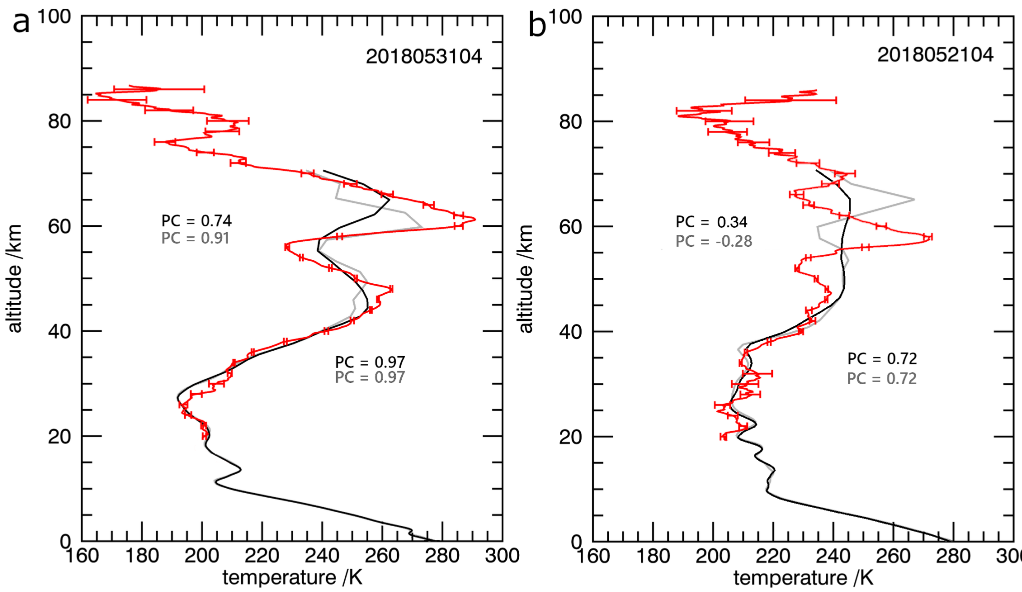

Figure 6. Example profiles for (a) 31 May 2018 04 UTC and (b) 21 May 2018 04 UTC of IFS temperature for the operational forecasts (black) and the experimental forecasts without the sponge (gray) and lidar temperature (red) with horizontal bars marking the uncertainty of the measurements. Corresponding Pearson correlation coefficients for 20 to 40 km and 40 to 65 km altitude are also given.

In Figure 7a, d, the PC for the mid-stratosphere (20 to 40 km altitude) is compared against the respective PC for upper stratosphere and lower mesosphere (40 to 65 km altitude) for May and August 2018. There is no general misrepresentation of upward propagating wintertime GWs in the middle atmosphere in the IFS. The PC for mid-stratosphere is larger than 0.7 for nearly all the profiles and phase deviations are mostly a feature of the upper stratosphere and lower mesosphere. There, the PC is often smaller than 0.7 (and sometimes even negative, i.e. antiphase). Phase deviations at higher altitudes are not related to the GW representation below in the mid-stratosphere. The GW generation and propagation processes in the IFS give a correct representation of the GW phases in the mid-stratosphere.

By grouping the profiles according to good ($PC > 0.7$) or poor ($PC \leq 0.7$) phase agreement in the upper stratosphere and lower mesosphere, it is possible to attribute the temperature differences to amplitude and phase errors. For the profiles with good phase agreement (Fig. 7b, e), the temperature differences are dominated by amplitude errors which can be large for individual profiles, but are often smaller than 10 K with a tendency for too cold temperatures in the upper stratosphere and lower mesosphere (at $T_{\text{lidar}} > 250$ K). Overestimated temperatures found for August at 40 km (Fig. 2) are related to amplitude errors (at $T_{\text{ECMWF}} = 230$ K in Figure 7e). For the profiles with poor phase agreement, the temperature differences (Fig. 7c, f) are a result of a combination of amplitude and phase errors. In the upper stratosphere and lower mesosphere ($T > 230$ K), the differences can clearly exceed 10 K in both directions, too warm (at $T_{\text{ECMWF}} = 250$ K in Figure 7c) or too cold (at $T_{\text{lidar}} > 250$ K in Figure 7c, f).

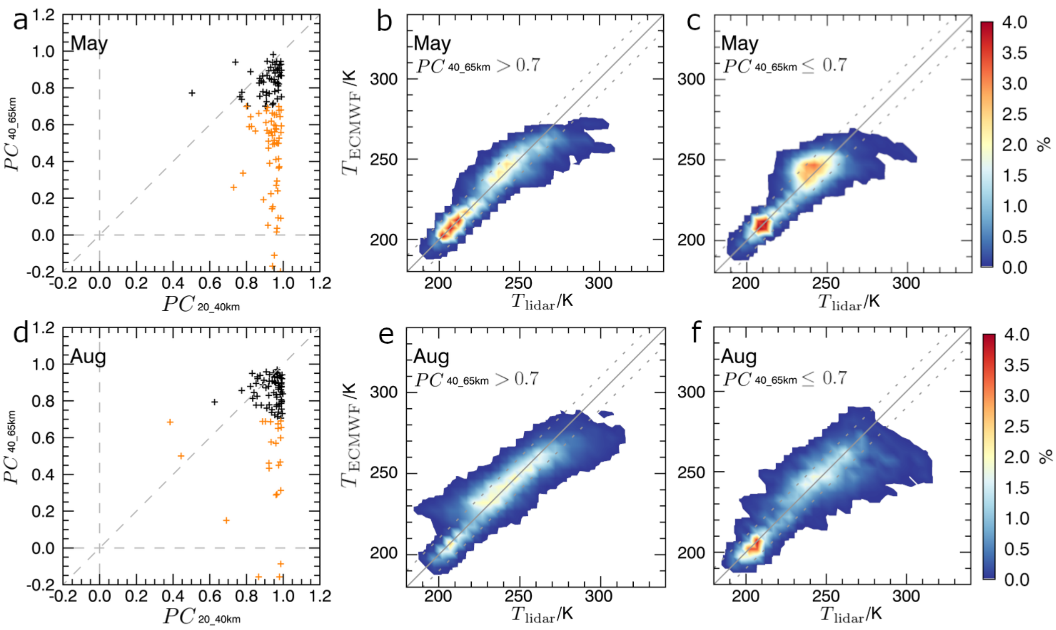

Figure 7. Pearson correlation coefficient calculated for lidar and IFS temperature data for the altitude ranges 20-40 km and 40-65 km (a, d). Orange color highlights low correlation (<0.7) at 40-65 km. Dashed gray lines mark the identity lines and $PC = 0$, respectively. Density distribution ($\Delta T = 5$ K) of data at all altitudes for lidar versus IFS temperatures for profiles showing high correlation (>0.7) at 40-65 km (b, e) and low correlation at 40-65 km (c, f). Solid gray lines mark the identity lines and the dotted lines the ±10 K range. Top panels are for May 2018 and bottom panels for August 2018.

In addition, Figure 7a, d shows that the amount of profiles per months that have poor phase agreement in the upper stratosphere and lower mesosphere is larger for May 2018 (orange crosses, representing 59 % of all profiles) compared to August 2018 (orange crosses, representing 24 % of all profiles). When considering the other months of the extended winter period (not shown), it becomes clear that phase deviations are also important for April 2018 (48 %) and slightly less important for June 2018 (38 %), July 2018 (33 %) and September 2018 (31 %). The amount of profiles that have poor phase agreement is also higher for May 2019 (46 %) and May 2020 (64 %) compared to July 2019 (24 %) and July 2020 (19 %). The amount of profiles that have a $PC < 0.7$ in the upper stratosphere and lower mesosphere is smaller in the extended summer period (0 % for Nov, Dec, Jan, Feb 2018 and 2 % and 4 % for Mar and Oct 2018, respectively) as expected due to less GW activity. Or in other words, high correlation ($PC > 0.7$) between the measurements and the IFS is found for around 95 % of the profiles above 40 km altitude in the extended summer period, which is consistent with the better agreement and smaller temperature differences in summer compared to winter (Fig. 2).

3.3 Gravity wave activity, intermittency, and effect of the model sponge

The energy of GWs in the lidar data and in the IFS is independent of the phase deviations and can be quantified by E_p (Eq. (6)). Figure 8 shows the annual cycle of $\overline{E_p}$ determined from lidar and IFS data for the altitude range of 35-50 km. The annual cycle with maximum (minimum) stratospheric GW activity in the winter (summer) that is characteristic for the Southern Andes region (Schroeder et al., 2009) is well reproduced by the IFS also above 35 km alitude. Coming back to the seasonal variability of the temperature differences between the lidar and the IFS, one finds that GW activity (Fig. 8) and the standard deviation of temperature differences (Fig. 5) show a similar annual cycle. The correlation coefficent for $\overline{E_p}$ of the lidar and $< T_\sigma >$ is 0.95 for 2018. The correlation is smaller (0.61) for $\overline{E_p}$ of the lidar and $< |\overline{T_{\mathrm{diff}}}| >$ which suggests that GW activity is less important for the monthly mean temperature profiles and their agreement between the lidar and the IFS.

Monthly mean $\overline{E_p}$ in the IFS is generally underestimated, which is a result of GW amplitude errors and corresponding underestimated T'. However, the reduction of $\overline{E_p}$ for July 2019 and 2020 compared to 2018 is reproduced by the IFS (see markers in Fig. 8). Besides $\overline{E_p}$, E_p of all the individual profiles vertically averaged for the same altitude range are also shown in Figure 8. This is to show that even though E_p is determined following Ehard et al. (2015) with T'^2 averaged along height (Tsuda et al., 2004), our E_p values are quantitatively very similar to the E_p values of Reichert et al. (2021) (see their Fig. 6). Moreover, E_p uncertainties due to lidar temperature uncertainties are insignificant at altitudes between 30 and 60 km (Reichert et al., 2021). E_p of the individual profiles also reveals that IFS indeed captures high E_p values of some strong GW events like the one in June 2018 (Fig. 8) that was analyzed in detail by N. Kaifler et al. (2020).

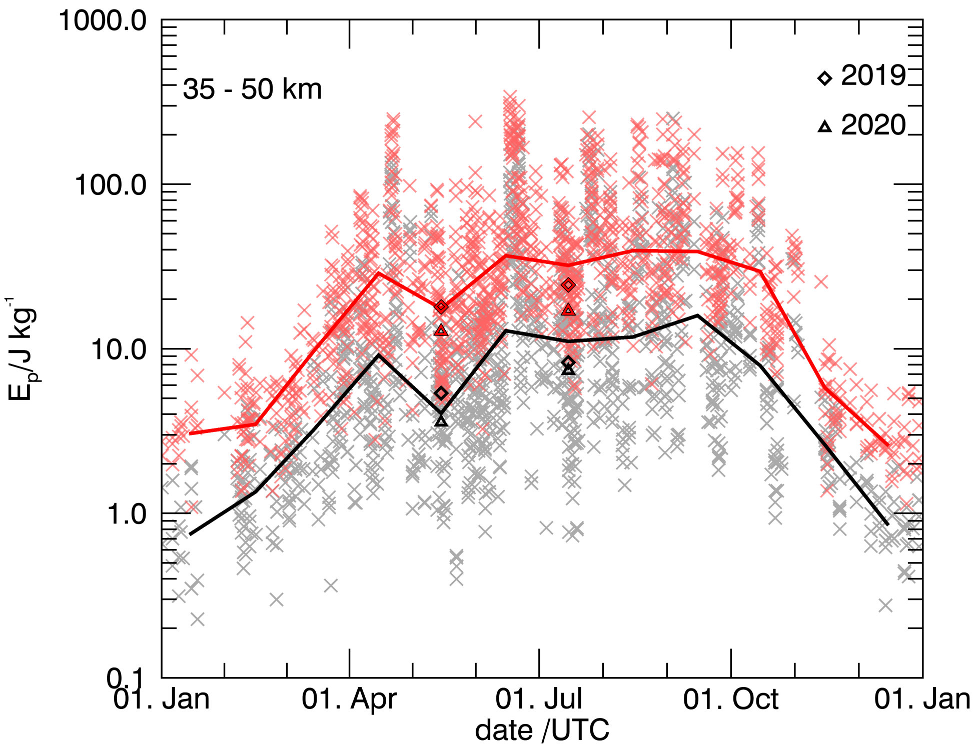

Figure 8. Annual cycle of $\overline{E_p}$ for the IFS (black) and for the lidar measurements (red) in the altitude range of 35 to 50 km for 2018. Diamonds and triangles show $\overline{E_p}$ for May and July 2019 and 2020, respectively. Crosses in the background show E_p of all the individual profiles in 2018 vertically averaged for the same altitude range.

The distributions of the E_p values for altitudes weakly affected by the model sponge (30-45 km) and strongly affected by the sponge (45-60 km) are shown in Figure 9a-d for May and August 2018. The distributions are in general log-normal with partly larger tails as can be seen in comparison with the probability density function computed from Eq. (8) using

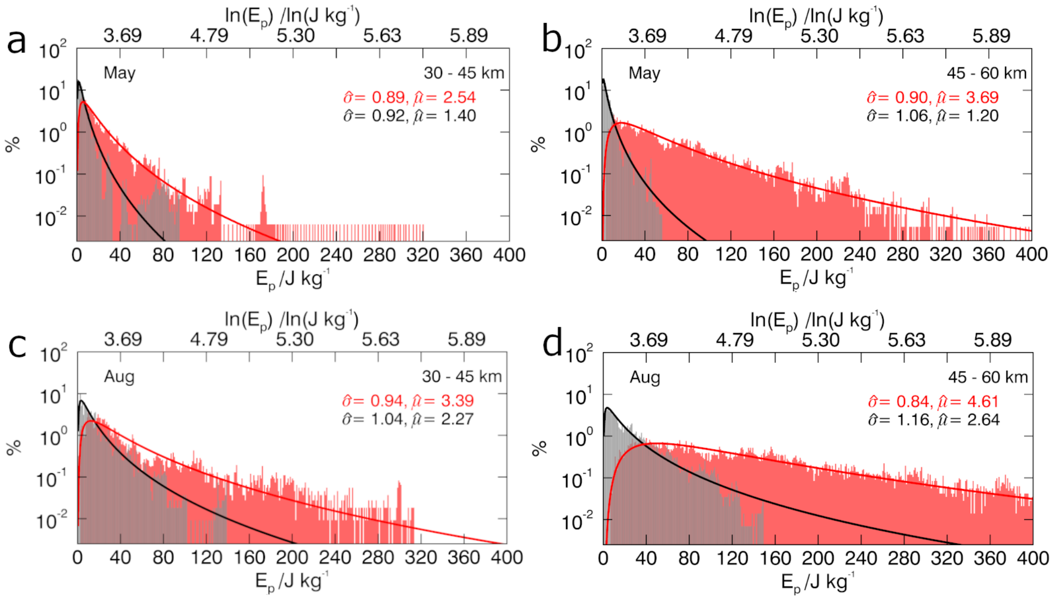

Figure 9. Distribution of E_p for the IFS operational forecasts (gray) and for the lidar measurements (light red) at an altitude range of 30 to 45 km (left) and 45 to 60 km (right) for May 2018 (top) and August 2018 (bottom). $\hat{\sigma}$ and $\hat{\mu}$ are the geometric standard deviation and expected value of the data distribution, respectively. Solid black and red lines show the probability density function of the log-normal distribution (Eq. (8)) computed with $\hat{\mu}$ and $\hat{\sigma}$.

$\hat{\mu}$ and $\hat{\sigma}$. Therefore, the expected or mean value $\hat{\mu}$ and the geometric standard deviation $\hat{\sigma}$ are better suited to describe the distributions than the arithmetic mean and standard deviation. $\hat{\sigma}$ of the distributions for the two months and for the lidar and the IFS data is close to unity but not exactly the same and clear differences are found for $\hat{\mu}$. Overall, GW activity is larger in August compared to May. $\hat{\mu}$ for the IFS is 55 to 66 % of $\hat{\mu}$ for the lidar measurements in the lower altitude range leading to $\overline{E_p}$ in IFS only reaching around 32 % of the $\overline{E_p}$ in the lidar measurements (Fig. 9a,c; Fig. 11). Nevertheless, the IFS captures some events of enhanced E_p as can be seen for example for May (E_p of 80 J/kg in Figure 9a).

In the upper altitude range, the comparison of the E_p distribution and the corresponding probability density function reveals that the IFS is missing the highest E_p values in the tail of the log-normal distribution, especially in August (Fig. 9b,d). $\overline{E_p}$ in IFS is only 8-14 % of the $\overline{E_p}$ of the lidar measurements (Fig. 9b,d; Fig. 11a,c). The no-sponge-runs show that this missing high E_p values and fairly low $\overline{E_p}$ are partly due to the sponge (Fig. 10b,d). The removal of the sponge leads to an increase of $\hat{\mu}$ and corresponding $\overline{E_p}$ to 25 % and

42 % of $\overline{E_p}$ of the lidar measurements for May and August 2018, respectively (Fig. 10b,d; Fig. 11b,d). Longer lead times of 25 to 35 hours further increase $\overline{E_p}$ in the no-sponge-runs to 33 % for May 2018 while it stays the same for August 2018 (not shown). At 30-45 km altitude, E_p remains similar when the sponge is removed with values generally smaller than 120 J/kg (Fig. 10a,c). Besides the effect of the sponge, small scale GWs that are not resolved in the IFS contribute to the larger E_p values of the lidar measurements. Regridding lidar temperature data to the 137 vertical levels of the IFS prior to the E_p calculations on the 100 m-grid eliminates GW structures from the lidar data that cannot be represented by the IFS simply because of the limited number of vertical levels. The high E_p values and averaged $\overline{E_p}$ of the lidar measurements are reduced by a similar amount as E_p values increase in the IFS when the sponge is removed (Fig. 10; Fig. 11). Clear differences between the E_p distributions of the original lidar data and the regridded lidar data can be seen for E_p values larger than 200 J kg^{-1} (240 J kg^{-1}) for May (August) at 45-60 km altitude (Fig. 10b,d; Fig. 10b,d). The contribution of scales unresolved in the IFS is likely even larger because this rough estimate does not consider the effective vertical resolution and scales not resolved horizontally.

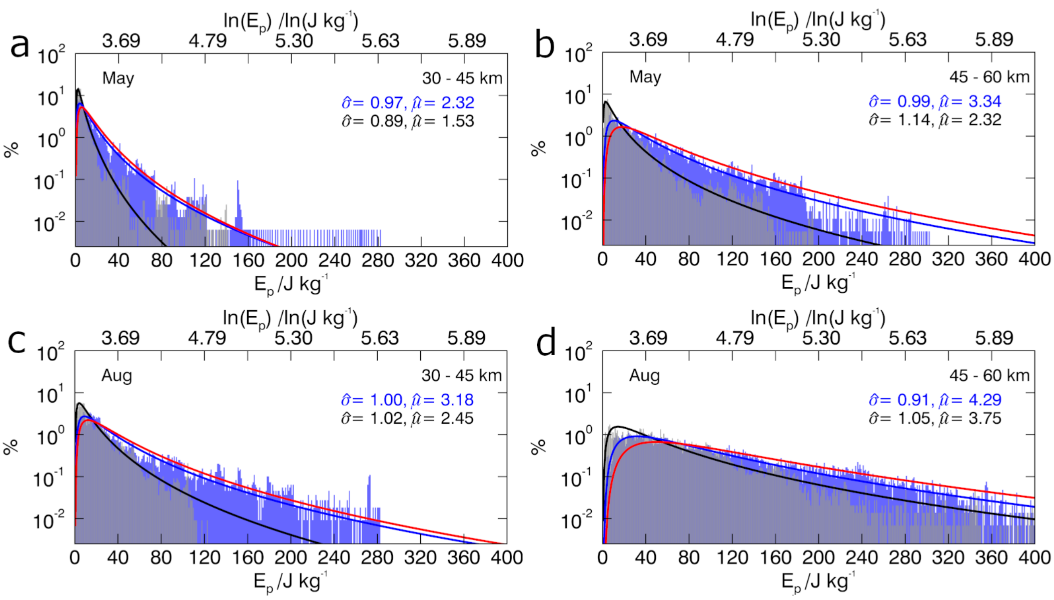

Figure 10. Same as Figure 9 but for the IFS experimental forecasts without the sponge (gray) and lidar data regridded to 137 vertical IFS levels prior to the analysis (light blue). Red line is from the original lidar data for direct comaprison (taken from Fig. 9).

To quantify the importance of extreme GW events that show large E_p in the analyzed full hour data compared to monthly $\overline{E_p}$ (i.e. intermittent GW activity), the Gini coefficient was calculated (Eq. (12)) for the two altitude regions for May and August 2018 (Tab. 2). Weaker extreme GW events in combination with smaller mean GW activity for May results in a similar Gini coefficient as for August when extreme GW events are stronger and the mean GW activity is larger. The lidar measurements and IFS agree in terms of GW intermittency at 30-45 km altitude. Above, the intermittency slightly decreases for the lidar measurements while it is almost constant for the IFS. The intermittency in the IFS slightly decreases (increases) with altitude for August (May) when the sponge is removed. The latter finding can be reproduced by repeating the analysis with better statistics for the full hourly data set of the IFS for May and August 2018, i.e. not limited to times where lidar observations are available.

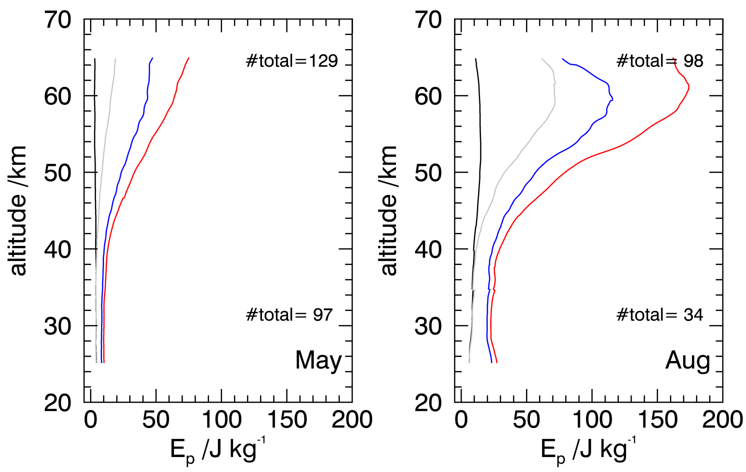

Figure 11. Monthly mean profiles of E_p for the operational forecasts (black), the experimential forecasts without the sponge (gray), the original lidar data (red), and the lidar data regridded to 137 vertical IFS levels prior to the analysis (blue) for May 2018 (left) and August 2018 (right). The number of profiles used for the statistics below (above) 30 km altitude is given at the bottom (top) part of the panels.

4 Discussion and Summary

Similar to previous studies for Europe (Le Pichon et al., 2015; Ehard et al., 2018; Marlton et al., 2021), we found a generally good agreement between the IFS and lidar temperature data up to 45 km altitude at higher mid-latitudes in the southern hemisphere in the

Table 2. Gini coefficient (Eq. (12)) for May and August 2018

data	month	30-45 km	45-60 km
CORAL	May 2018	0.51	0.46
IFS	May 2018	0.51	0.52
IFS no sponge	May 2018	0.52	0.57
CORAL	Aug 2018	0.51	0.42
IFS	Aug 2018	0.52	0.51
IFS no sponge	Aug 2018	0.49	0.45

lee of the Southern Andes. Monthly mean differences are smaller than 2 K between 20 and 40 km altitude for all months and usually negative, i.e. temperatures are underestimated. An exeption was August 2018, when monthly mean differences are +2 K at 40 km altitude. Higher up around the stratopause between 45 km and 55 km altitude, which is above the peak altitude of assimilated radiances (2 hPa) in the IFS and influenced by the strong sponge, conditions are more variable in time and the sign of the monthly mean temperature deviations changes throughout the year. The largest positive monthly mean difference (4 K) occurs in May 2018. The largest negative monthly mean difference of -10 K related to the warm stratopause (approx. 268 K) is found in August 2018, July 2019, and July 2020. This means the underestimation of upper stratospheric temperatures at Rio Grande in winter lies within the range of mean deviations that were found for the older IFS cycle 41r1 (-8 K) and cycle 41r2 (-20 K) in the northern hemisphere for December 2015 (Ehard et al., 2018). For the extended summer period (October to March 2018), the monthly mean differences are smaller than -4 K between 45 km and 55 km altitude. For these months, differences for individual profiles are rarely found outside the range of ±5 K. The standard deviation of the temperature differences and maximum differences for individual profiles are significantly larger and increase with altitude in the winter (> ±10 K). In addition, high correlation ($PC > 0.7$) between the measurements and the IFS is found for most temperature profiles (around 95 %) above 40 km altitude in the extended summer period. The better agreement between the IFS and the lidar measurements in the summer months previously found for the northern hemisphere (Le Pichon et al., 2015) also manifests for the southern hemisphere and a more recent cycle of the IFS. The high correlation between the annual cycle of the temperature deviations and of the stratospheric GW activity supports the hypothesis that the seasonal variability of the temperature deviations over Rio Grande is related to the stratospheric GW activity.

The analysis of individual profiles for May and August 2018, revealed that the PC gives high correlation (>0.7) only when the GWs of measurements and IFS are in phase. This is the case at altitudes below 40 km which means that the temperature differences at these altitudes are mainly due to deviations in amplitudes and not due to phase deviations. Phase deviations are found to be a feature of the upper stratosphere and lower mesosphere and are therefore likely a result of the propagation and representation of GWs in the middle atmosphere in the IFS. Resulting temperature differences at these altitudes are as such a combination of amplitude and phase deviations. The enhanced amount of profiles that show poor phase agreement for May (> 40 %) compared to July to September (< 33 %) could be the reason why satellite observations in the upper stratosphere are rejected by the 4D-Var in the IFS more frequently in May. A quantitative evaluation of phase deviations in the wintertime temperature profiles that are shaped by GWs has not been published before for the IFS according to our knowledge. For an eight-day period with strong GW activity in June 2018, N. Kaifler et al. (2020) found good agreement in amplitude and phase of the mountain waves in lidar and IFS data over Rio Grande. Such information can only be extracted when instantaneous temperature profiles are available instead of nightly means (e.g., Le Pichon et al., 2015) and when the analysis is not only restricted to monthly mean statistics (e.g., Ehard et al., 2018).

The analysis of the annual cycle of GW activity in the middle and upper stratosphere complements the findings by Schroeder et al. (2009) for the Andes and reveals that the IFS captures well the winter maximum and summer minimum also at altitudes above 30 km. In general, the IFS underestimates E_p in the middle atmosphere over Rio Grande and the discrepancy is increasing with altitude. $\overline{E_p}$ of the IFS above 45 km altitude is only around 10 % of $\overline{E_p}$ derived from the lidar observations. Similar results are found for ERA5 in Strelnikova et al. (2021) who show that GW potential energy densities of ERA5 at 55 km altitude are on average one order of magnitude smaller (i.e. 10 %) when compared to two European lidar stations. However, there can be quite good agreement below 45 km altitude for individual events like the one at Rio Grande in June 2018 analyzed in detail by N. Kaifler et al. (2020). While the removal of the sponge can lead to increasing temperature differences at certain altitudes for profiles with phase deviations, it has a positive effect on E_p (i.e., an increase) above 45 km altitude in IFS because E_p is independent of the actual phase of the GWs. $\overline{E_p}$ increases from only ≈10 % of the $\overline{E_p}$ of the lidar measurements to 25 % and 42 % for May and August 2018, respectively, when the sponge is removed. This shows that the sponge is an important but not the only cause for a reduced $\overline{E_p}$ in the IFS. Unresolved GWs and model resolution are also important, in particular the vertical resolution that is coarse in the upper stratosphere and lower mesosphere.

GW intermittency is usually calculated for GW momentum fluxes that were determined for example from balloon (Plougonven et al., 2013), satellite (Wright et al., 2013; Hindley et al., 2019) or radar (Minamihara et al., 2020) measurements. Because it is not reasonable to directly compare GW intermittency by means of the Gini coefficient for GW momentum fluxes and E_p of various observations that are sensitive to different parts of the GW spectrum and that are focusing on different periods of time without fundamental evaluations, the discussion here is limited to relative changes of the Gini coefficient with altitude over Rio Grande in the lidar measurements and the IFS. GW intermittency slightly decreases for the lidar measurements from 30-45 km to 45-60 km altitude. It is almost constant for the operational IFS data but slightly decreases (increases) with altitude for August (May) when the sponge is removed in the model. Generally, it is interesting that the GW intermittency in the IFS is close to the intermittency in lidar measurements, even though the E_p distributions of the IFS are shifted to smaller E_p values compared to the distributions of the lidar measurements. In regions where orographic GWs dominate, the intermittency decreases with height when GWs with large momentum flux are removed at altitudes where the background wind matches the ground-based phase velocity of the GWs (Minamihara et al., 2020). However, this mechanism cannot explain the steep decline of GW intermittency found around the tropopause in the PANSY MST radar data at Syowa station, Antarctica, and partial reflection due to discontinuities in static stability (i.e. the tropopause in this case), that can result in a more continuous distribution of GWs in the vertical, is mentioned as one possible explanation (Minamihara et al., 2020). Changing static stability in the vicinity of the stratopause at around 50 km (Fig. 1) can likely have a similar effect on the GW intermittency in the middle atmosphere over Rio Grande. In addition, large-amplitude orographic GWs can break or dissipate well below their critical level at the mesopause in winter or propagate horizontally out of the observational volume of the ground-based lidar (Ehard et al., 2017). All these processes are potentially important and could lead to decreasing intermittency with altitude at the location of Rio Grande.

In summary, this study presents the first detailed quantitification of local deviations between middle atmosphere lidar temperature measurements and IFS temperatures for the stratospheric GW hot spot region of the Southern Andes. It was found that the ability of the IFS in most accurately representing temperatures of the middle atmosphere over Rio Grande depends on the altitude range and the season. In particular, smoother conditions in summer are better captured by current model configurations than the more complex wintertime conditions involving GWs. The shortcomings in the representation of middle atmosphere GWs in the IFS are characterized by amplitude and phase deviations that contribute to the site-specific temperature deviations. While amplitude deviations in the IFS are due to the sponge and unresolved GWs, the origin of the shift in the phases of the GWs in the IFS compared to the lidar measurements, that is often observed in the upper stratosphere and lower mesosphere, still needs to be found. In the mid-stratosphere, the GW generation and propagation processes in the IFS give a correct representation of the GW phase. Investigating this topic in more detail could help to understand why phase deviations are happening most frequently in fall, i.e. May, and improving phase representation could help to prevent the rejection of satellite observations in the IFS data assimilation system. Moreover, improving amplitude deviations in the upper stratosphere and lower mesosphere by e.g., a weaker sponge will only work if GW phases are represented correctly.

Acknowledgments

Part of this research was conducted within programs of the German Science foundation under Grants GW-TP/DO1020/9-1 and PACOG/RA1400/6-1 in the frame of the DFG research group MS-GWAVES (FOR1898) and of the German Federal Ministry for Education and Research under Grant 01LG1907 (project WASCLIM) in the framework of ROMIC (Role of the Middle Atmosphere in Climate). BK and NK thank Jose-Luis Hormaechea and Alejandro de la Torre for support with logistics and operation of the lidar at Estacion Astronomica Rio Grande (EARG) station in Rio Grande. SG and IP thank the ISSI Team "New Quantitative Constraints on OGW Stress and Drag" at the ISSI Bern for fruitfull discussions of related results. A 2-months visit of SG at ECMWF in collaboration with IP and IS was funded by ECMWF. Access to ECMWF operational forecast data was granted through the special project "Deep Vertical Propagation of Internal Gravity Waves".

Data availability The ECMWF IFS and CORAL temperature profile data used in the study are available at the HALO-DB (https://halo-db.pa.op.dlr.de/mission/111; license

CC BY 4.0 and ECMWF's Terms of Use apply). Dataset numbers (XXXX) are:7905-7925 and directly accessible via https://halo-db.pa.op.dlr.de/dataset/XXXX.

References

Alexander, M. J. (1998). Interpretations of observed climatological patterns in stratospheric gravity wave variance. *J. Geophys. Res. Atmos.*, *103*(D8), 8627–8640. doi: 10.1029/97JD03325

Baldwin, M. P., & Dunkerton, T. J. (2001). Stratospheric harbingers of anomalous weather regimes. *Science*, *294*(5542), 581–584. doi: 10.1126/science.1063315

Baumgaertner, A. J. G., & McDonald, A. J. (2007). A gravity wave climatology for antarctica compiled from challenging minisatellite payload/global positioning system (champ/gps) radio occultations. *Journal of Geophysical Research: Atmospheres*, *112*(D5). doi: 10.1029/2006JD007504

Byrne, N. J., & Shepherd, T. G. (2018). Seasonal persistence of circulation anomalies in the southern hemisphere stratosphere and its implications for the troposphere. *Journal of Climate*, *31*(9), 3467 - 3483. doi: 10.1175/JCLI-D-17-0557.1

Eckermann, S. D., Ma, J., Hoppel, K. W., Kuhl, D. D., Allen, D. R., Doyle, J. A., ... Love, P. T. (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). *Monthly Weather Review*, *146*(8), 2639 - 2666. doi: 10.1175/MWR-D-17-0386.1

ECMWF. (2003). Part III: Dynamics and Numerical Procedures. In P. White (Ed.), *IFS Documentation CY23R4*. Retrieved from `https://www.ecmwf.int/node/13272`

Ehard, B., Kaifler, B., Dörnbrack, A., Preusse, P., Eckermann, S. D., Bramberger, M., ... Rapp, M. (2017). Horizontal propagation of large-amplitude mountain waves into the polar night jet. *J. Geophys. Res. Atmos.*, *122*(3), 1423–1436. doi: 10.1002/2016JD025621

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

Ehard, B., Malardel, S., Dörnbrack, A., Kaifler, B., Kaifler, N., & Wedi, N. (2018). Comparing ECMWF high-resolution analyses with lidar temperature measurements in the middle atmosphere. *Q.J.R. Meteorol. Soc.*, *144*(712), 633-640. doi: 10.1002/qj.3206

Ern, M., Preusse, P., Alexander, M. J., & Warner, C. D. (2004). Absolute values of gravity wave momentum flux derived from satellite data. *Journal of Geophysical Research: Atmospheres*, *109*(D20). doi: 10.1029/2004JD004752

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

Garfinkel, C. I., & Hartmann, D. L. (2011). The Influence of the Quasi-Biennial Oscillation on the Troposphere in Winter in a Hierarchy of Models. Part I: Simplified Dry GCMs. *Journal of the Atmospheric Sciences*, *68*(6), 1273 - 1289. doi: 10.1175/2011JAS3665.1

Gisinger, S., Dörnbrack, A., Matthias, V., Doyle, J. D., Eckermann, S. D., Ehard, B., ... Rapp, M. (2017). Atmospheric conditions during the deep propagating gravity wave experiment (DEEPWAVE). *Mon. Weath. Rev.*, *145*(10), 4249-4275. doi: 10.1175/MWR-D-16-0435.1

Hertzog, A., Alexander, M. J., & Plougonven, R. (2012, 11). On the Intermittency of Gravity Wave Momentum Flux in the Stratosphere. *Journal of the Atmospheric Sciences*, *69*(11), 3433-3448. doi: 10.1175/JAS-D-12-09.1

Hindley, N. P., Wright, C. J., Smith, N. D., Hoffmann, L., Holt, L. A., Alexander, M. J., ... Mitchell, N. J. (2019). Gravity waves in the winter stratosphere over the southern ocean: high-resolution satellite observations and 3-d spectral analysis. *Atmospheric Chemistry and Physics*, *19*(24), 15377–15414. doi: 10.5194/acp-19-15377-2019

Hoffmann, L., Xue, X., & Alexander, M. (2013). A global view of stratospheric gravity wave hotspots located with atmospheric infrared sounder observations. *J. Geophys. Res. Atmos.*, *118*(2), 416–434.

Kaifler, B., & Kaifler, N. (2021). A Compact Rayleigh Autonomous Lidar (CORAL) for the middle atmosphere. *Atmospheric Measurement Techniques*, *14*(2), 1715–1732. doi: 10.5194/amt-14-1715-2021

Kaifler, N., Kaifler, B., Dörnbrack, A., Rapp, M., Hormaechea, J., & Torre, A. (2020, 09). Lidar observations of large-amplitude mountain waves in the stratosphere above tierra del fuego, argentina. *Scientific Reports*, *10*, 14529. doi: 10.1038/s41598-020-71443-7

Khaykin, S. M., Hauchecorne, A., Wing, R., Keckhut, P., Godin-Beekmann, S., Porteneuve, J., ... Schmitt, J. (2020). Doppler lidar at observatoire de haute-provence for wind profiling up to 75 km altitude: performance evaluation and observations. *Atmospheric Measurement Techniques*, *13*(3), 1501–1516. doi: 10.5194/amt-13-1501-2020

Le Pichon, A., Assink, J. D., Heinrich, P., Blanc, E., Charlton-Perez, A., Lee, C. F., … Claud, C. (2015). Comparison of co-located independent ground-based middle atmospheric wind and temperature measurements with numerical weather prediction models. *Journal of Geophysical Research: Atmospheres*, *120*(16), 8318-8331. doi: https://doi.org/10.1002/2015JD023273

Marlton, G., Charlton-Perez, A., Harrison, G., Polichtchouk, I., Hauchecorne, A., Keckhut, P., … Steinbrecht, W. (2021). Using a network of temperature lidars to identify temperature biases in the upper stratosphere in ECMWF reanalyses. *Atmospheric Chemistry and Physics*, *21*(8), 6079–6092. doi: 10.5194/acp-21-6079-2021

Minamihara, Y., Sato, K., & Tsutsumi, M. (2020). Intermittency of gravity waves in the antarctic troposphere and lower stratosphere revealed by the pansy radar observation. *Journal of Geophysical Research: Atmospheres*, *125*(15), e2020JD032543. doi: https://doi.org/10.1029/2020JD032543

Plougonven, R., Hertzog, A., & Guez, L. (2013). Gravity waves over antarctica and the southern ocean: consistent momentum fluxes in mesoscale simulations and stratospheric balloon observations. *Quarterly Journal of the Royal Meteorological Society*, *139*(670), 101-118. doi: 10.1002/qj.1965

Polichtchouk, I., Hogan, R., Shepherd, T., Bechtold, P., Stockdale, T., Malardel, S., … Magnusson, L. (2017). What influences the middle atmosphere circulation in the IFS? (809). doi: 10.21957/mfsnfv15o

Preusse, P., Eckermann, S. D., & Ern, M. (2008). Transparency of the atmosphere to short horizontal wavelength gravity waves. *Journal of Geophysical Research: Atmospheres*, *113*(D24). doi: https://doi.org/10.1029/2007JD009682

Preusse, P., Ern, M., Bechtold, P., Eckermann, S. D., Kalisch, S., Trinh, Q. T., & Riese, M. (2014). Characteristics of gravity waves resolved by ECMWF. *Atmospheric Chemistry and Physics*, *14*(19), 10483–10508. doi: 10.5194/acp-14-10483-2014

Reichert, R., Kaifler, B., Kaifler, N., Dörnbrack, A., Rapp, M., & Hormaechea, J. L. (2021). High-Cadence Lidar Observations of Middle Atmospheric Temperature and Gravity Waves at the Southern Andes Hot Spot. *J. Geophys. Res. Atmos.*. (accepted) doi: 10.1029/2021JD034683

Rennie, M., & Isaksen, L. (2020). The NWP impact of Aeolus Level-2B Winds at ECMWF. (864). doi: 10.21957/alift7mhr

Rüfenacht, R., Baumgarten, G., Hildebrand, J., Schranz, F., Matthias, V., Stober, G., …

Kämpfer, N. (2018). Intercomparison of middle-atmospheric wind in observations and models. *Atmospheric Measurement Techniques*, *11*(4), 1971–1987. doi: 10.5194/amt-11-1971-2018

Schroeder, S., Preusse, P., Ern, M., & Riese, M. (2009). Gravity waves resolved in ECMWF and measured by SABER. *Geophysical Research Letters*, *36*(10). doi: https://doi.org/10.1029/2008GL037054

Shepherd, T. G., Polichtchouk, I., Hogan, R., & Simmons, A. (2018, 06). Report on stratosphere task force. (824). doi: 10.21957/0vkp0t1xx

Shepherd, T. G., Semeniuk, K., & Koshyk, J. N. (1996). Sponge layer feedbacks in middle-atmosphere models. *Journal of Geophysical Research: Atmospheres*, *101*(D18), 23447-23464. doi: https://doi.org/10.1029/96JD01994

Simmons, A., Soci, C., Nicolas, J., Bell, B., Berrisford, P., Dragani, R., ... Schepers, D. (2020). Global stratospheric temperature bias and other stratospheric aspects of ERA5 and ERA5.1. (859). doi: 10.21957/rcxqfmg0

Strelnikova, I., Almowafy, M., Baumgarten, G., Baumgarten, K., Ern, M., Gerding, M., & Lübken, F.-J. (2021). Seasonal cycle of gravity wave potential energy densities from lidar and satellite observations at 54° and 69°n. *Journal of the Atmospheric Sciences*, *78*(4), 1359 - 1386. doi: 10.1175/JAS-D-20-0247.1

Tsuda, T., Ratnam, M. V., May, P. T., Alexander, M. J., Vincent, R. A., & MacKinnon, A. (2004). Characteristics of gravity waves with short vertical wavelengths observed with radiosonde and gps occultation during dawex (darwin area wave experiment). *J. Geophys. Res. Atmos.*, *109*(D20).

Wright, C. J., Osprey, S. M., & Gille, J. C. (2013). Global observations of gravity wave intermittency and its impact on the observed momentum flux morphology. *Journal of Geophysical Research: Atmospheres*, *118*(19), 10,980-10,993. doi: https://doi.org/10.1002/jgrd.50869

Yamashita, C., Liu, H.-L., & Chu, X. (2010). Gravity wave variations during the 2009 stratospheric sudden warming as revealed by ecmwf-t799 and observations. *Geophysical Research Letters*, *37*(22). doi: https://doi.org/10.1029/2010GL045437