Final and cumulative response to the reviewers' comments on our manuscript

"An intercomparison of stratospheric gravity wave potential energy densities from METOP GPS-radio occultation measurements and ECMWF model data"

by M. Rapp, A. Dörnbrack, and B. Kaifler

submitted for publication in Atmospheric Measurement Techniques (amt-2017-346)

Weßling, 22.12.2017

We greatly appreciate the reviewers' positive assessment of our work as well as their constructive comments. For the revised version, all comments have been addressed. In the following we respond to all comments point by point. For clarity, we first repeat the comments of the referees (in black), we then respond to these (in blue), and then indicate the changes made to the text where appropriate (in red).

Reviewer 1

[..] Overall, the paper shows that METOP GPS-radio occultations are a promising dataset for gravity wave analysis. The paper is very well written, and publication in AMT is recommended after addressing my minor comments.

We are grateful to the reviewer for this encouraging judgment.

My main comment is that some more discussion is needed regarding the separation of temperature altitude profiles into background and temperature fluctuations due to gravity waves. The selected method is a vertical filter with cutoff at 15 km vertical wavelength. The main idea is that all variations with wavelengths shorter than 15 km are assumed to be gravity waves. This, however, is not stated clearly enough, and the pros and cons of this approach should be briefly discussed as detailed in my specific comments.

This is an excellent point raised by both reviewers. As a consequence, we have revised the text in several places as indicated below in response to the more specific points. In addition, mainly in response to reviewer 2, we have added an additional figure (Figure 15) in the revised manuscript in which we compare latitude-altitude distributions of monthly mean zonal mean fields of E_P derived with the standard method applied in this study (i.e., using a fifth order Butterworth filter with a cutoff at 15 km wavelength in the vertical) as well as derived using a horizontal background determination method. Similarities and differences from both analysis techniques are discussed and a recommendation for future work is formulated (see also our detailed answer to comment 6 below).

Specific comments

(1) about Sect. 2.2: Please clarify that the spatial resolution mentioned in this section corresponds to the horizontal grid spacing. Atmospheric waves that are resolved by these data sets have scales that are considerably longer. According to Skamarock (2004) only scales exceeding the grid spacing by several times are resolved. Skamarock, W. C.: Evaluating Mesoscale NWP Models Using Kinetic Energy Spectra, Mon. Wea. Rev., 132, 3019-3032, 2004.

Thanks for pointing this out. We have corrected the text accordingly.

[..] These have a horizontal grid spacing of 16 km (T_L1279) and were evaluated on 25 pressure levels between 1000 and 1 hPa which we converted to geopotential heights and interpolated them on a regular vertical grid with 1 km spacing. We note that according to Skamarock (2004) only scales exceeding the grid spacing by several times are resolved.

The horizontal grid spacing of the data set is approximately 80 km (T255).

(2) p.5, l.12, Sect. 2.3 Here, you call the reduced scatter of RO-wet temperatures a reduced "uncertainty range". Is this justified, or are RO-wet temperatures too smooth? Above 30km RO-wet temperatures show a reduced scatter with respect to ECMWF IFS. However, at these altitudes the influence of a priori data should be increasing, and it is known that at high altitudes ECMWF is known to suffer from hyper-diffusion. Could it therefore happen that temperature fluctuations are suppressed in RO-wet temperatures because of an increasing influence of relatively smooth a priori data?

The reviewer again has a good point here. In order to give a more neutral description of Figure 4 we have changed the wording “uncertainty range” to “variability range”.

(3) p.5, l.28 The idea behind using a Butterworth filter should be stated more clearly, and shortcomings mentioned. As far as I understand, variations in the vertical with scales longer than 15km are assumed to be the "background" (climatological structure plus planetary waves), while shorter scales are assumed to be fluctuations due to atmospheric gravity waves. This separation of scales will work well with a few exceptions. One exception is the tropopause region, which has been discussed in detail in the current paper. Another exception is the tropical stratosphere. While vertical wavelengths of planetary waves in the extratropics are quite long, this is different in the tropics where Kelvin waves usually have vertical wavelengths that are comparable to those of gravity waves. See also (5).

(4) p.5, l.28 Please mention that the use of the Butterworth filter in vertical direction has the advantage of being applicable in the same manner to all data sets considered, thus allowing a fair comparison.

(5) p.6 l.10-14 Epot close to the equator will be high-biased due to Kelvin waves Kelvin waves in the tropics can have quite short vertical wavelengths, comparable to those of gravity waves. Kelvin waves can have considerable temperature variances of a few K^2 on zonal average, and corresponding zonal average values of Epot could easily reach values of around 5 J/kg, which is non-negligible. This is particularly important because Fig.6 represents a case of tropical easterlies. Under these conditions the amplitudes of Kelvin waves will be amplified. A climatology Kelvin waves in the tropical stratosphere is given, for example, in Ern et al. (2008) Ern, M., Preusse, P., Krebsbach, M., Mlynczak, M. G., and Russell III, J. M.: Equatorial wave analysis from SABER and ECMWF temperatures, Atmos. Chem. Phys., 8, 845- 869, 2008.

Reply to 3, 4, and 5: We agree that it is critical to better explain the idea behind using a Butterworth filter. Also the limitations of this approach in the tropics must be more clearly mentioned and it needs to be pointed out more directly that the most important advantage of this approach is that all four data sets (including the ground based lidar data) can be analyzed with identical analysis routines. The text has been changed as follows:

For this study, we follow the approach of Ehard et al. (2015), i.e., we apply a fifth-order Butterworth filter with a cutoff wavelength of 15 km to vertical temperature profiles from the RO-measurements, from ERA Interim, from the IFS, as well as from ground based lidar -measurements. Applying this filter to altitude profiles implies that scales longer than 15 km are assumed to be the "background" (climatological structure plus planetary waves), denoted $T_0(z)$, while shorter scales are assumed to be fluctuations due to atmospheric gravity waves. This separation is expected to work well except for the tropical stratosphere where Kelvin waves are known to occur with vertical wavelengths well below 15 km (e.g., Ern et al., 2008; Randell and Wu, 2005). Hence, E_P must be expected to be biased high in the tropics.
Nevertheless we stick to this approach since it has the advantage that all data sets analyzed in this study can be treated with identical analysis routines thus allowing us to directly and quantitatively compare E_P-values from four independent data sets.

(6) p.9, about the Epot correction: Do you think that monthly average temperatures are sufficient for deriving a correction, as proposed? Or may there be changes in the background on shorter time scales that would require averaging over shorter time intervals? Of course, this may be beyond the scope of the current paper, but should be carefully considered before making this correction operational. See also p.11, l.16+17

We agree with both reviewers that this correction indeed warrants a more in-depth discussion. In the new Figure 15 we show a monthly mean zonal mean distribution of E_P that was derived using a horizontal background determination method. In the same figure we further show monthly mean zonal mean fields of vertical kinetic energies, i.e., $VE=0.5*w^2$, due to gravity waves (see e.g., Geller and

Gong, 2010). Note that the mean vertical velocity is zero so that non-zero VE-structures are mainly due to gravity waves. This comparison clearly shows that E_P-values derived with the horizontal background derivation agree much better in terms of their morphology to vertical kinetic energy values than E_P-values determined from vertical profiles (compare Figures 14 and 15). Figure 15 further shows that the correction method improves the qualitative comparison between E_P and VE but that it cannot eliminate all features that are apparently not due to gravity waves. Closer inspection of the data sets reveals that this is partly because some of the non-gravity wave structures (mainly the TTIL) are not zonally homogeneous such that correcting for them using zonal mean fields cannot eliminate the corresponding signatures completely.

In the context of Figure 15, the following text has been added:

We finally attempt to determine the quality of the corrected Ep -values in Figure 14 by comparing them to Ep -values using a horizontal background determination method. Horizontal estimation of T_0 was previously found to be superior to a vertical background determination by Khaykin (2016) and Schmidt et al. (2016). While the sampling statistics of the METOP RO-data on a daily basis (i.e., only 1100 profiles distributed over the whole globe) is too poor to allow us to apply a horizontal background determination to them we may easily perform a corresponding analysis of the high resolution IFS-data. For this purpose, the spectral model output of the IFS for December 2015 has been reconstructed at T42, i.e., at a horizontal grid spacing of 500 km. These fields have then been used as background temperatures $T_0(z,\lambda,\phi)$, where λ is latitude and ϕ is longitude, in order to compute monthly mean zonal mean distributions of Ep . Such monthly mean zonal mean Ep -distributions for December 2015 are presented in Figure 15. In the same figure we also show corresponding fields of the vertical kinetic energy, $VE = 0.5*w^2$ (Geller and Gong, 2010). Note that VE is a good indicator for gravity waves in the stratosphere since vertical velocities due to other air motions are significantly smaller. While VE-values are significantly smaller than Ep -values (by about a factor of 1000 in the IFS-model) it is still instructive to compare the spatial morphology of the corresponding fields. This comparison clearly reveals that the proposed correction of Ep -distributions derived using a vertical background determination (see Figure 14 and related text) improves the comparison between Ep and VE but that it cannot eliminate all features that are apparently not due to gravity waves. Closer inspection of the data sets reveals that this is partly because some of the non-gravity wave structures (mainly the TTIL) are not zonally homogeneous such that correcting for them using zonal mean fields cannot eliminate the non-gravity wave structures completely. We hence conclude that this correction may be recommended for application to data sets that can only be analyzed using a vertical background determination method such as for the METOP data with relatively scarce sampling statistics. However, even after this correction, regions within +/- 30° latitude around the equator need to be considered with care due to additional potential contamination of Ep by Kelvin waves or other planetary scale features. In any case, if the sampling statistics allows, our analysis clearly shows that in general a horizontal background determination is advantageous in that it better avoids contributions to E_P that are not caused by gravity waves.

Likewise, statements in abstract and conclusions were added.

Abstract: This correction may be recommended for application to data sets that can only be analyzed using a vertical background determination method such as the METOP data with relatively scarce sampling statistics. However, if the sampling statistics allows, our analysis also shows that in general a horizontal background determination is advantageous in that it better avoids contributions to E_P that are not caused by gravity waves.

Conclusions: In addition, this technique to derive and correct E_P based on vertical profiles was compared to an alternative method applying a horizontal background temperature determination method to IFS-data. We find that the above introduced correction may be recommended for application to data sets that can only be analyzed using a vertical background determination method such as the METOP data with relatively scarce sampling statistics. However, if the sampling statistics allows, our analysis also shows that in general a horizontal background determination is advantageous in that it better avoids contributions to E_P that are not caused by gravity waves like the TTIL and potentially also Kelvin waves and other planetary scale features with short vertical wavelengths (i.e., less than 15 km).

(7) p.11, l.22 Please include the information that Kelvin waves could produce a high bias of Epot in the tropics.

Done.

TECHNICAL COMMENTS:
p.2, l.7 GW are a major means to couple the -> GW are an important mechanism that couples the

Changed as suggested.

p.2, l.30 please correct: ECMWF = "European Centre for Medium-Range Weather Forecasts"

Thanks. Changed as suggested.

Fig.1a Is the red dot at 90N an artifact, or is this a real accumulation of ROs?

It is indeed an artifact and has been removed.

p.3, l.18 The expression in parentheses is confusing; suggestion a corresponding gridding (i.e., 36 x 36 5deg latitude x 10deg longitude bins) -> a corresponding gridding of 36 x 36 grid points (i.e., 5deg latitude x 10deg longitude bins)

Thanks. Changed as suggested.

p.4, l.8 Please check whether it is correct that T1279 corresponds to a horizontal resolution of 8km.To my knowledge, the ECMWF grid uses 2560 points at the equator, corresponding to 15km grid spacing. The numbers of T255/80km that given for ERA-Interim should however be correct.

The reviewer is correct. We have corrected the grid spacing to 16 km.

p.4, l.12 from -> starting from

We corrected the typo.

p.7, l.23 wavelengths if the phase fronts are perpendicular to the line of sight, -> wavelengths than is the case if the phase fronts are perpendicular to the line of sight,

Thanks. Changed as suggested.

p.8, l.33 cleat -> clear

Corrected.

p.13, l.33 stratopsphere -> stratosphere

Corrected.

p.13, l.34: Ern et al., 2016 -> Ern et al., 2017

Corrected.

p.14, l.12: LOVE -> Love ??

Corrected.

p.14, l.15 Hei, H., T., T. T., and Hirooka: -> Hei, H., Tsuda, T., and Hirooka, T.:

Corrected. Thanks for the careful reading!

Reviewer 2

[..]The interpretation of the data sets is well set in the scientific background of previous publications and largely sound. A few exceptions are listed below. These points need to be corrected. The paper is generally well to read and recommended for publication in AMT after some revisions.

Thanks very much for the careful review and the encouraging judgement of our work.

General remarks.

1) Choice of the useful vertical range for GPS-RO.
The choice of lower boundary of 20 km is conditioned by potential aliasing of tropopause layer's structure and variability as GW-induced perturbations. Although this excludes the entire extra-tropical lower stratosphere from the analysis, it allows to consider and compare the global distributions of GW Ep without heavy stipulations regarding the effects of tropical tropopause. My concern here mostly regards the upper boundary of 40 km. GPS-RO is a powerful tool for temperature profiling, which is why a good knowledge of actual capacities of this technique is crucial for atmospheric community. The majority of RO-based GW studies restrict the analysis to altitudes below 35 km because the uncertainty and the noise become too large at high altitudes. While this is confirmed by the results of this paper, I would love to see among the conclusions some more definitive statements regarding the usefulness of z>35 km RO data for GW retrieval.

At the end of Section 3 we had already written: "For this reason, we will exclude altitudes below 20 km from our analysis and focus on the altitude range between 20 and 40 km only, knowing, of course, that the largest altitudes need to be treated with care since noise of RO-data is known to pick up significantly above 35 km altitude (Marquardt and Healy, 2005)."

We do agree with the reviewer that our analyses in Section 4 do confirm that the actual useful range of data is below 35 km. This has now also been explicitly pointed out in the conclusions.

Also, it is well known that noise in RO-data picks up substantially above 35 km such that several previous studies have recommended to restrict the useful range of RO-data for GW analysis to below 35 km (e.g., Schmidt et al., 2008). This previous recommendation is clearly supported by our analysis.

2) Choice of Ep derivation method (Sect. 3).
As mentioned in the paper, there is a number of techniques for isolating the GW-induced perturbations in temperature profiles from the background atmospheric state. Here the authors opted to use a vertical detrending method based on a Butterworth filter. The advantage of the vertical detrending is that it can be applied to both local and global observations, enabling direct RO-lidar comparison. At that, I wonder how different would the results be for RO and ECMWF (particularly the TTIL issue) had the authors used a horizontal detrending method for GW Ep retrieval, which seems to better handle the lower stratosphere region (Schmidt et al., 2016; Khaykin et al., 2016). Indeed, when the authors subtract the zonal-mean Ep profiles (which is already some sort of horizontal detrending), the TTIL anomaly disappears. A recommendation to use this correction is put forth in the conclusions but I wonder, wouldn't it be better just to use one of the horizontal detrending methods instead? I believe the authors should better justify the choice of Ep derivation method in consideration of its shortcomings before recommending it for future use.

As already pointed out in our response to reviewer 1, this is an excellent and important comment. In order to address it, we have pointed out more clearly in the revised manuscript that we are using a vertical detrending method in order to treat all data sets with the same analysis procedure such that derived EP-values are directly comparable. The text added in the revised manuscript in response of this comment as well as to comment 5 of reviewer 1 is as follows:

For this study, we follow the approach of Ehard et al. (2015), i.e., we apply a fifth-order Butterworth filter with a cutoff wavelength of 15 km to vertical temperature profiles from the RO-measurements, from ERA Interim, from the IFS, as well as from ground based lidar measurements. [..] we stick to this approach since it has the advantage that all data sets analyzed in this study can be treated with identical analysis routines thus allowing us to directly and quantitatively compare E_P-values from four independent data sets.

In addition, we have added Figure 15 with its corresponding discussion (repeated below from our response to reviewer 1):

We finally attempt to determine the quality of the corrected Ep -values in Figure 14 by comparing them to Ep -values using a horizontal background determination method. Horizontal estimation of T_0 was previously found to be superior to a vertical background determination by Khaykin (2016) and Schmidt et al. (2016). While the sampling statistics of the METOP RO-data on a daily basis (i.e., only 1100 profiles distributed over the whole globe) is too poor to allow us to apply a horizontal background determination to them we may easily perform a corresponding analysis of the high resolution IFS-data. For this purpose, the spectral model output of the IFS for December 2015 has been reconstructed at T42, i.e., at a horizontal grid spacing of 500 km. These fields have then been used as background temperatures $T_0(z,\lambda,\phi)$, where λ is latitude and ϕ is longitude, in order to compute monthly mean zonal mean distributions of Ep . Such monthly mean zonal mean Ep -distributions for December 2015 are presented in Figure 15. In the same figure we also show corresponding fields of the vertical kinetic energy, $VE = 0.5*w^2$ (Geller and Gong, 2010). Note that VE is a good indicator for gravity waves in the stratosphere since vertical velocities due to other air motions are significantly smaller. While VE-values are significantly smaller than Ep -values (by about a factor of 1000 in the IFS-model) it is still instructive to compare the spatial morphology of the corresponding fields. This comparison clearly reveals that the proposed correction of Ep -distributions derived using a vertical background determination (see Figure 14 and related text) improves the comparison between Ep and VE but that it cannot eliminate all features that are apparently not due to gravity waves. Closer inspection of the data sets reveals that this is partly because some of the non-gravity wave structures (mainly the TTIL) are not zonally homogeneous such that correcting for them using zonal mean fields cannot eliminate the non-gravity wave structures completely. We hence conclude that this correction may be recommended for application to data sets that can only be analyzed using a vertical background determination method such as for the METOP data with relatively scarce sampling statistics. However, even after this correction, regions within +/- 30° latitude around the equator need to be considered with care due to additional potential contamination of Ep by Kelvin waves or other planetary scale features. In any case, if the sampling statistics allows, our analysis clearly shows that in general a horizontal background determination is advantageous in that it better avoids contributions to E_P that are not caused by gravity waves.

Likewise, statements in abstract and conclusions were added.

Abstract: This correction may be recommended for application to data sets that can only be analyzed using a vertical background determination method such as the METOP data with relatively scarce sampling statistics. However, if the sampling statistics allows, our analysis also shows that in general a horizontal background determination is advantageous in that it better avoids contributions to E_P that are not caused by gravity waves.

Conclusions: In addition, this technique to derive and correct E_P based on vertical profiles was compared to an alternative method applying a horizontal background temperature determination method to IFS-data. We find that the above introduced correction may be recommended for application to data sets that can only be analyzed using a vertical background determination method such as the METOP data with relatively scarce sampling statistics. However, if the sampling statistics allows, our analysis also shows that in general a horizontal background determination is advantageous in that it better avoids contributions to E_P that are not caused by gravity waves like the TTIL and potentially also Kelvin waves and other planetary scale features with short vertical wavelengths (i.e., less than 15 km).

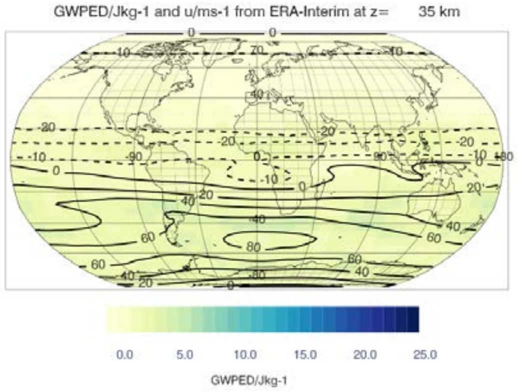
Specific remarks.
P4, L17-19. Although the correct references are in place, the fact that the compared data sets are not independent requires some more specific information on the assimilation of RO data in ECMWF IFS and ERA-Interim.

We have now pointed out more clearly that the data sets are not completely independent.

Thus, ECMWF-model fields and METOP RO-data are obviously not completely independent.

P6, L20-22. It is indeed surprising that ERA doesn't see the Scandinavian GW hotspot. The validity of the proposed explanation (invoking coarser resolution of ERA) could be verified by examining global Ep distribution for June, when the strong Patagonian GW hot spot is well pronounced.

This is a very good suggestion. As suggested we have inspected the E_P-distributions for June 2015 in the Southern hemisphere based on both ERA and IFS-data. While IFS-data do show moderate EP-values over the Patagonian hot spot at 35 km altitude (see lower right panel in Figure 12), ERA-data don't as shown below.



This supports our hypothesis that the coarser resolution of ERA-data leads to the inability to reproduce localized mountain wave activity. A corresponding short statement has been added to the text.

Note that we have checked this interpretation by also comparing Ep -distributions over the well-known Patagonian GW-hot spot for June 2015. While METOP and IFS-data show clear signatures of strong GW activity in this region (see Figure 12), ERA-Interim again misses to reproduce this GW activity (not shown).

P.7, L17-31. What is missing in this discussion is the mention of the large difference in vertical resolution of GPS-RO and lidar at high altitudes. P7, L33-34. The sentence could be reformulated to make it clearer that it is the derived Ep values that are low-biased and not the actual temperature data.

The vertical resolution of the two techniques is actually not so much different: 900m for the lidar and ~1.5 km for the RO-data (Kaifler et al., 2005; Kursinski et al., 1997). This information has been added to the text. Also the sentence has been reformulated to make clear that Ep is low-biased and not T.

P8, L27. The relatively large bias is rather seen below 23 km.

This is correct and has been changed in the text.

P.9,L.10-11. What is somewhat controversial here is that the higher N2 values within the TTIL derived from RO should lead to lower Ep, whereas the results suggest the opposite. It is understandable that RO should better resolve the fluctuations, but invoking the differences in N2 field in this context should be done more carefully.

Here, one may not confuse the N^2-values shown in Figure 13 and the N^2-values used in the computation of Ep. For Figure 13, the N^2-values were calculated from monthly mean zonal mean T-profiles thus containing information on the climatological temperature profile. For Ep, however, N^2-values are calculated from the filtered individual temperature profiles such that scales smaller than 15 km (such as the TTIL) are suppressed. Hence the N^2-values shown in Figure 13 are an illustration of the climatological small scale structure in the temperature profile but do not enter the Ep-calculation. This has now been clarified in the text.

Note that the N^2-values in Figure 13 were computed from monthly mean zonal mean-temperatures. These must not be confused with the N^2-values used in our EP -calculation which is based on T_0-profiles. Remember that T_0-profiles result from filtering individual temperature profiles with a 5th-order

Butterworth-filter with cutoff wavelength at 15 km such that T_0-profiles only contain spatial scales larger than 15 km, and hence do not contain information on the TTIL.

Technical remarks.

5,L9. Remove “km” after the right parenthesis

Done.

P.8, L.31. “At these altitudes” => at the level of lowest correlation?

We changed the wording to “at the altitude levels of lowest correlation”

P.9, L24. “larger altitudes” => higher altitudes? Throughout Sect. 6, comma is erroneously used as a decimal separator instead of a point.

We changed the wording to “higher altitudes”. Also, we consistently replaced the commas by points whenever used as a decimal separator throughout Section 6.

Fig. 3. The data are missing in the left-hand panels.

This must have been a problem of the reviewer’s pdf-file or pdf viewer. The one we see on the AMT-website does show the data.

Fig. 3 and 9. The X-axis of right-hand panels could be reduced to enhance the readability of the histograms.

Done as suggested.

Fig. 4 left panel. The X-axis scale could be reduced to say 0.8 – 1.1

Done as suggested. We reduced the scale to 0.9 – 1.1

Fig. 5 upper panels. X-axis caption should be T and not T’

Thanks, this has been corrected.

Fig. 6. The black-shaded land areas whenever Ep values are too low are somewhat confusing.

We actually found this presentation less confusing than showing the continental contours by mere black lines. Since this appears to be a matter of taste we have left this figure as it was.

An intercomparison of stratospheric gravity wave potential energy densities from METOP GPS-radio occultation measurements and ECMWF model data

Markus Rapp[1,2], Andreas Dörnbrack[1], and Bernd Kaifler[1]

[1]Deutsches Zentrum für Luft- und Raumfahrt, Institut für Physik der Atmosphäre, Oberpfaffenhofen, Germany
[2]Meteorologisches Institut München, Ludwig-Maximilians-Universität München, Munich, Germany

Correspondence to: Markus Rapp (markus.rapp@dlr.de)

Abstract. Temperature profiles based on radio occultation (RO) measurements with the operational European METOP-satellites are used to derive monthly mean global distributions of stratospheric (20 - 40 km) gravity wave (GW) potential energy densities (E_P) for the period July 2014 - December 2016. In order to test whether the sampling and data quality of this data set is sufficient for scientific analysis we investigate to which degree the METOP-observations agree quantitatively with ECMWF operational analysis (IFS-data) and reanalysis (ERA-Interim) data. A systematic comparison between corresponding monthly mean temperature fields determined for a latitude-longitude-altitude grid of 5° by 10° by 1 km is carried out. This yields very low systematic differences between RO and model data below 30 km (i.e., median temperature differences is between -0,2 and +0,3 K) which increases with height to yield median differences of +1,0 K at 34 km and +2,2 K at 40 km. Comparing E_P-values for three selected locations at which also ground based lidar measurements are available yields excellent agreement between RO and IFS-data below 35 km. ERA-Interim underestimates E_P under conditions of strong local mountain wave forcing over Norther Scandinavia which is apparently not resolved by the model. Above 35 km, RO-values are consistently much larger than model values which is likely caused by the model sponge layer which damps small scale fluctuations above ~32 km altitude. Another reason is the well known significant increase of noise in RO-measurements above 35 km. The comparison between RO and lidar data reveals very good qualitative agreement in terms of the seasonal variation of E_P, however, RO-values are consistently smaller than lidar values by about a factor of two. This discrepancy is likely caused by the very different sampling characteristics of RO and lidar observations. Direct comparison of the global data set of RO and model E_P-fields shows large correlation coefficients (0.4 - 1.0) with a general degradation with increasing altitude. Concerning absolute differences between observed and modelled E_P-values, the median difference is relatively small at all altitudes (but increasing with altitude) with an exception between 20 and 25 km where the median difference between RO- and model-data is increased and where also the corresponding variability is found to be very large. The reason for this is identified as an artifact of the E_P-algorithm: this erroneously interprets the pronounced climatological feature of the tropical tropopause inversion layer (TTIL) as GW activity hence yielding very large E_P-values in this area and also large differences between model and observations. This is because the RO-data show a more pronounced TTIL than IFS and ERA-Interim. We suggest a correction for this effect based on an estimate of this ‘artificial’ E_P using monthly mean zonal mean temperature profiles. **This correction may be recommended for application to data sets that can only be analyzed using a vertical background determination method such as the**

METOP data with relatively scarce sampling statistics. However, if the sampling statistics allows, our analysis also shows that in general a horizontal background determination is advantageous in that it better avoids contributions to E_P that are not caused by gravity waves.

1 Introduction

It has long been known that momentum and energy transport by gravity waves (henceforth abbreviated as GW) is of major importance for the mean thermal and dynamical state of the middle atmosphere (Lindzen, 1981; Holton and Alexander, 2000). Being mainly excited in the troposphere by flow over terrain, by convection, or by spontaneous emission, GW may propagate both vertically and horizontally over large distances to deposit their momentum and energy upon instability or transience far away from their source (e.g., Fritts and Alexander, 2003; Sato et al., 2009; Preusse et al., 2009; Sato et al., 2012; Bölöni et al., 2016). Thus, GW are an important mechanism that couples the middle and upper atmosphere to the troposphere (e.g., Lübken et al., 2010, and references therein). In addition, it has recently been shown that GW also couple the middle atmosphere downward to the troposphere (Kidston et al., 2015, and references therein). With minimum horizontal scales as small as 10 km GW must still be parameterized in global climate models with typical horizontal resolutions of a few hundred kilometers. Hence, the development of physics-based parameterizations of GW and their effect on the mean flow have been identified as a major research focus in the climate research community (Shepherd, 2014).

Given this large importance of GW, it is not surprising that efforts have been undertaken and are under way trying to characterize GW sources, their propagation, as well as their dissipation and wave-mean flow interaction with complementary experimental, theoretical and numerical techniques (see, e.g., Fritts and Alexander, 2003; Plougonven and Zhang, 2014; Fritts et al., 2016; Wagner et al., 2017; Sutherland, 2010; Nappo, 2012, for recent reviews, overview papers, and text books). Ground based remote sensing with lidars and radars and in-situ observations with balloons, research aircraft, and sounding rockets are critically important for process studies. However, global satellite observations are needed to determine dominant tropospheric source regions and processes as well as global propagation pathways and the resulting gravity wave drag imposed on the mean flow to constrain GW parameterizations for climate and weather prediction models (Alexander et al., 2010; Geller et al., 2013). Since the pioneering work by Fetzer and Gille (1994),Wu and Waters (1996), and Eckermann and Preusse (1999) there have been many attempts to characterize the global distribution of gravity wave activity using such different remote-sensing techniques as Limb (e.g., Ern et al., 2004; Preusse et al., 2009; Ern et al., 2011; Zhang et al., 2012) and Nadir sounders (e.g., Hoffmann et al., 2016; Ern et al., 2017), as well as GPS-based radio occultation measurements (e.g., Tsuda et al., 2000; Hei et al., 2008; Schmidt et al., 2008; Fröhlich et al., 2007; Hindley et al., 2015; Šácha et al., 2015; Khaykin et al., 2015; Khaykin, 2016; Schmidt et al., 2016).

This paper focusses on the derivation of gravity wave potential energy densities (E_P) from GPS radio occultation (RO) measurements onboard the operational METOP-A and METOP-B-satellites operated by EUMETSAT (=European Organisation for the Exploitation of Meteorological Satellites) and the subsequent systematic comparison of E_P-fields with ECMWF (European Centre for Medium-Range Weather Forecasts) operational forecast and reanalysis data. This is done to answer the question whether the sampling and data quality of the two operational METOP-satellites is sufficient to characterize the global stratospheric gravity wave activity (measured in terms of E_P) on a monthly basis. Furthermore, we investigate whether the METOP-observations agree quantitatively with the ECMWF model-fields such that the latter can be used for the interpretation of observational results. Accordingly, this paper is organized as follows: in Section 2 we describe the data base of METOP-A and METOP-B radio occultation temperature data obtained between July 2014 and December 2016. In addition, we give a brief introduction to the ECMWF data sets used for comparison with the RO-data. We compare both temperature data sets (RO and ECMWF-data) as a baseline for the subsequent comparison of derived E_P-values. In Section 3 we describe our approach to derive E_P followed by Section 4 where we thoroughly compare RO E_P-data to corresponding ECMWF-data sets. Similarities and differences are discussed in Section 5 in which we will also derive and discuss a correction for erroneous interpretation of the tropical tropopause inversion layer (TTIL) as gravity wave activity. Finally, the major findings of this study are summarized in Section 6 in which also suggestions for future work will be made.

2 Data base

2.1 METOP-A/B GPS radio occultation data

The METOP-A and -B satellites orbit the Earth in a polar low Earth orbit and are the platforms for a variety of instruments supporting the European Weather Services including the Global Navigation Satellite System Receiver for Atmospheric Sounding (GRAS) with which GPS radio occultation measurements are performed delivering tropospheric humidity and tropospheric and stratospheric temperature profiles. During typical months, these two satellites record a total of $\sim$ 35.000 - 40.000 radio occultations. A typical sampling pattern in terms of the latitude and longitude distribution of the number of RO per month is shown in Figure 1. This sampling is determined by the orbital geometry of the METOP-satellites on the one hand and the GPS-satellites on the other. Figure 1 reveals that there are typically between 10 and 50 occultations per 5 degree latitude and 10 degree longitude interval with maximum sampling at latitudes between 20 - 60° North and South and minima near the poles and at the equator. Note that we will use a corresponding gridding of 36 $\times$ 36 grid points (i.e., 5° latitude by 10° longitude bins) throughout this entire study.

The METOP RO-data are provided by the Radio Occultation Meteorology Satellite Application Facility (ROM SAF) on an operational basis in near real time and can be downloaded from www.romsaf.org. The primary measured quantity is the bending angle of the GPS radio waves as they transverse the refracting atmosphere. From bending angle profiles corresponding refractivity profiles can be derived from which in turn also temperature profiles can be determined (Kursinski et al., 1997). The latter can be done by either assuming that the refraction is entirely due to dry air (resulting in so-called ‘dry’ temperatures) or by accounting for tropospheric water vapor by using additional information, e.g., from operational numerical weather forecast data in the framework of a one dimensional variational algorithm that uses ECMWF IFS data as a priori information (ROM SAF, 2014a, b, and references therein). The latter approach is pursued by the ROM SAF and corresponding temperature data are denoted ‘wet’ temperatures. For the current study we will mainly use dry instead of wet temperatures since the latter have been derived using model output and might hence not be considered as ‘pure’ measurements. Nevertheless, we will also briefly consider wet temperatures and compare them to the more ‘original’ dry ones. Note that the ROM SAF provides dry and wet temperatures from July 2014 onwards only. Hence, in this study we restrict ourselves to the period from July 2014 to December 2016, i.e., a total of 30 months of data.

METOP temperature profiles are provided on geopotential heights which will be used here as the vertical coordinate. The fundamental vertical resolution of the technique, Δz, is limited by diffraction as the GPS rays pass through the atmosphere and results in about Δz=1 - 1.4 km in the altitude range between 15 and 40 km. Over this vertical interval, the horizontal line-of-sight resolution can be estimated to be around 190 - 270 km due to the limb geometry of the observations (see Kursinski et al., 1997; Hindley et al., 2015, for details).

2.2 ECMWF operational analysis and reanalysis data

For comparison to the METOP RO-data we use two different data sets provided by the ECMWF: one is the operational analyses from the Integrated Forecast System (IFS). **These have a horizontal grid spacing of about 16 km (T_L1279) and were evaluated on 25 pressure levels between 1000 and 1 hPa which we converted to geopotential heights and interpolated them on a regular vertical grid with 1 km spacing. We note that according to Skamarock (2004) only scales exceeding the grid spacing by several times are resolved.** Model output is available every six hours. Details about the model can be found in Malardel and Wedi (2016) and in references therein.

The second model data set that we use is the ERA-Interim reanalysis. ERA-Interim is a global atmospheric reanalysis starting from 1979 which is based on a 2006 release of the IFS. **The horizontal grid spacing of the data set is approximately 80 km (T255).** For the current study, model fields were evaluated on 37 standard pressure levels between 1000 and 1hPa which we converted to geopotential heights and interpolated them on a regular vertical grid with 1 km spacing. For details about ERA-Interim see Dee et al. (2011).

Please note that ECMWF does assimilate RO-bending angle data (among many other data sets) from a variety of instruments including (but not limited to) the METOP-data for both the ERA Interim reanalysis as well as for the IFS analyses (see Poli et al., 2010, as well as the ECMWF web site). **Thus, ECMWF-model fields and METOP RO-data are obviously not completely independent.**

2.3 Comparison between RO and ECMWF-temperature data

In this subsection we systematically compare RO temperatures with ERA-Interim and IFS model data. As a start, Figure 2 shows zonal mean temperatures for the months March, June, and December 2015 derived from METOP GPS RO-dry data (left column), GPS-RO wet data (middle column), and from ERA Interim. Note that from now on, we will refer to METOP GPS-RO-dry and -wet data as ‘RO-dry’ and ‘RO-wet’ data for brevity. Overall, all data sets agree well with notable differences, however, between the dry temperatures and the other two data sets in the troposphere and at the highest altitudes above 40 km. These findings are not surprising given that the retrieval for wet temperatures uses ECMWF IFS data as a priori information, that the assumption of dry conditions is certainly violated in the troposphere, and that the quality of RO observations in general decreases significantly above $\sim$ 40 km altitude (Marquardt and Healy, 2005). In the following, we hence restrict our comparison to altitudes between 20 - 40 km.

For a more quantitative comparison, we have binned the ECMWF data sets on the same space and time grid as the RO-data, i.e., mean profiles were determined for the period of one month and a latitude-longitude-altitude grid of 5° by 10° by 1 km. Figure 3 shows corresponding scatter plots between RO dry temperatures and corresponding IFS-data for all 30 months of data considered in this study (i.e., July 2014 - December 2016) as well as histograms of the temperature differences between the data for three selected altitudes. This reveals very large correlation coefficients close to one between the data with a general degradation of the (still very good) correlation as well as an increasing bias between the data with increasing altitude. The full altitude variation of the correlation coefficients between the considered data sets as well as the median temperature differences along with corresponding 10% and 90% percentiles is shown in Figure 4. This again shows an almost perfect correlation between ERA-Interim and IFS-data (as expected) as well as between the RO-wet temperatures and the IFS. Again, only the dry temperatures show a notable disagreement from the other data sets at altitudes above $\sim$ 35 km. This is further quantified with the median biases (and percentiles) shown in the right panel of Figure 4 which shows a median bias of +1 K (+2 K) between RO-dry temperatures and the IFS (i.e., IFS temperatures are larger than dry RO temperatures) at 34 km (40 km) with a corresponding large **variability range** as indicated by the percentiles. We note that these values are in excellent quantitative agreement with a previous study in which GPS RO-observations were compared to ECMWF-data (Scherllin-Pirscher et al., 2011). Compared to this bias of the RO-dry data, it is again not surprising to see that the RO-wet temperatures show a much smaller bias (close to zero) to the IFS and that also the corresponding **variability range** is much reduced. Both derived biases and **variability ranges** agree well with previous findings of an analysis of the ROM SAF as described in the corresponding validation report (ROM SAF, 2014c). In all, RO-temperatures agree well with ERA-Interim and IFS-temperatures such that it appears justified to proceed and next compare corresponding E_P-values.

3 Derivation of E_P

We next turn to the derivation of E_P from the various input temperature data sets considered in this study. E_P is defined as follows:

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

where g is acceleration of gravity, $N^2 = \frac{g}{T_0}\cdot(\frac{dT_0}{dz}+\frac{g}{c_p})$ is the (squared) bouyancy-frequency with the specific heat capacity of air for constant pressure c_p, $T^{'}$ is the temperature perturbation owing to the GW, and T_0 is the background temperature. The overbar denotes averaging, which is here carried out over the spatial domain of the latitude-longitude-grid and the time period of one month. In Equation 1 all quantities depend on height z except for g for which we use a constant value of 9.81 ms^{-2}. The main challenge in deriving $E_p(z)$ from measured temperature profiles lies in the separation between background and perturbations. Different studies have used various approaches such as filtering of profiles in the vertical or in the horizontal provided that the horizontal sampling is sufficient. See Khaykin (2016) and Ehard et al. (2015) for recent critical discussions of the advantages and disadvantages of different techniques.

For this study, we follow the approach of Ehard et al. (2015), i.e., we apply a fifth-order Butterworth filter with a cutoff wavelength of 15 km to vertical temperature profiles from the RO-measurements, from ERA Interim, from the IFS, as well as from ground based lidar-measurements. Applying this filter to altitude profiles implies that scales longer than 15 km are assumed to be the "background" (climatological structure plus planetary waves), denoted $T_0(z)$, while shorter scales are assumed to be fluctuations due to atmospheric gravity waves. This separation is expected to work well except for the tropical stratosphere where Kelvin waves are known to occur with vertical wavelengths well below 15 km (e.g., Ern et al., 2008; Randel and Wu, 2005). Hence, E_P must be expected to be biased high in the tropics. Nevertheless we stick to this approach since it has the advantage that all data sets analyzed in this study can be treated with identical analysis routines thus allowing us to directly and quantitatively compare E_P-values from four independent data sets.

Resulting $T_0(z)$-profiles are then used to derive $N^2(z)$-profiles. Arbitrarily chosen sample profiles from RO-dry data are shown in Figure 5. Figure 5 shows both cases with strong (middle panel) and weak GW-activity (left panel). These sample profiles further show that the background temperature determination has weaknesses in cases with a very pronounced tropopause as in the right panel. We will come back to this issue in more detail in Section 5. Here, neither the pronounced tropopause (at around 17 km), nor the inversion layer above (i.e., between 20 - 25 km) is well captured by the Butterworth filter resulting in unrealistically large temperature perturbations which might not be confused with real gravity wave-induced temperature perturbations. This is a general problem with all techniques analyzing vertical temperature profiles which has motivated many authors to exclude the tropopause region and the lowest altitudes above it from further analyses (see e.g. Schmidt et al., 2008, for a detailed discussion and an approach to derive GW properties in the vicinity of the tropopause). For this reason, we will exclude altitudes below 20 km from our analysis and focus on the altitude range between 20 and 40 km only, knowing, of course, that the largest altitudes need to be treated with care since noise of RO-data is known to pick up significantly above ~35 km altitude (Marquardt and Healy, 2005).

4 Comparison of METOP E_P-values with ECMWF model data and ground based lidar measurements

We next present a systematic comparison of E_P-values derived from METOP RO-dry temperatures, from the IFS, and from ERA-Interim. As an initial impression, Figure 6 shows monthly mean latitude-longitude cross sections of E_p at selected altitudes of 30, 33, 36, and 39 km for December 2015, respectively. At 30 km, the RO-data reveal pronounced GW activity over Scandinavia, over the Iberian peninsula and North Africa, and in a band in the vicinity of the equator with strongest activity in the tropical central Pacific (135 - 180° E). Moving to 33 km altitude, E_P-values increase with pronounced activity still over Scandinavia, strong activity at around 40°N in the Atlantic storm track region, and an additional activity center over the northern part of South America. At larger altitudes, these general features remain, but become smeared out geographically. Generally speaking, this overall morphology of GW activity is well reproduced by both the IFS and ERA-Interim with some notable differences. First of all, E_p-values from the IFS and ERA-Interim are generally smaller than corresponding RO-values, with this discrepancy increasing with increasing altitude. Secondly, ERA-Interim does not capture the GW activity over Scandinavia that is clearly seen in the RO-data and also in the IFS-data. The latter finding is likely due to the significantly coarser horizontal resolution (and hence also a coarser resolution of orography) which keeps the ERA-Interim reanalysis from capturing rather localized orographic gravity wave activity as the one here seen over Scandinavia. **Note that we have checked this interpretation by also comparing E_P-distributions over the well known Patagonian GW-hot spot for June 2015. While METOP and IFS-data show clear signatures of smoderate GW activity in this region (see Figure 12), ERA-Interim again misses to reproduce this GW activity (not shown).**

For a more detailed comparison, we have next extracted time series of E_P for different altitude bands at selected locations. We have selected three locations at which we have conducted extended ground based Rayleigh lidar observations of E_P and are hence in a position not only to compare RO-data with the two different model data sets but also with the ground based data. These locations are: Lauder, New Zealand (45° S, 169.7° E), where ground based Rayleigh lidar measurements were conducted from June through December 2014 (Kaifler et al., 2015; Fritts et al., 2016), Sodankylä, Finnland (67° N, 26° E), where observations were taken from September 2015 until May 2016 (Kaifler et al., 2017), and finally in the German Bavarian Forest (48.8° N, 13.7° E) with measurements from May until December 2016. Figure 7 shows time series of monthly mean E_P from July 2014 through December 2016 for these locations and for the altitude ranges 15-25 km, 25-35 km, and 35-45 km, respectively. Note that we have binned the model data to the same latitude-longitude grid as the RO-data for a proper comparison. Figure 7 overall reveals a very good fit between RO and model data: in the altitude range from 15-25 km, RO and model data fit very well both in terms of absolute values as well as in terms of month-to-month variation for all three locations (top panels). At altitudes between 25-35 km (middle panels), the agreement is still very good, however, peak RO values are underestimated by both models. This is particularly pronounced for Sodankylä (left), where local mountain wave activity is likely causing the strong wintertime E_P-peak. Consistent with the results shown in Figure 6 this peak is qualitatively well reproduced (but still slightly underestimated) by the IFS but completely missed by ERA-Interim due to the much coarser horizontal resolution of the latter. Finally, at the highest altitudes, the overall seasonal variation of E_P that is observed with the RO-sensors is reproduced by the models, but modelled E_P-values are smaller than those derived from RO-observations by factors between 2 and 3. This is as expected since the sponge layer in the ECMWF-models starts strongly damping any small scale structures above 10 hPa or $\sim$32 km (Jablonowski and Williamson, 2011; Ehard, 2017) and since RO-measurement noise is picking up substantially above 35 km (see Marquardt and Healy, 2005, and our analysis in Figure 4 and corresponding discussion).

Next, we compare the same RO time series to local E_P-observations obtained with Rayleigh-lidar (see Figure 8). The portable lidar systems as well as the data analysis procedure used during the three campaigns have been described in detail in Kaifler et al. (2015) and Kaifler et al. (2017), respectively. In short, Rayleigh lidar measurements yield relative density profiles at altitudes where pure molecular scatter accounts for the signal, i.e., from above the stratospheric aerosol layer. Hence, data are available for altitudes above $\sim$30 km (and below $\sim$90 km) but may be extended to lower altitudes after careful analysis making sure that stratospheric aerosol-scatter did not contribute to the signal. Relative density profiles are then converted to temperatures applying hydrostatic downward integration. Finally, E_P-values are derived in the same manner as for the RO-data described above (see Section 3).

The comparison shown in Figure 8 reveals that lidar and RO-data generally show very similar seasonal variation. However, the comparison also shows that the local lidar observations yield significantly larger E_P-values by up to a factor of $\sim$2. This is likely because the lidar observations are sensitive to a larger part of the gravity wave spectrum than the RO-observations. As described in Section 2.1, the horizontal line of sight of RO-observation is approximately 190 - 270 km. Hence, depending on the orientation of the wave vector relative to this line of sight, the RO-technique may not resolve waves with horizontal wavelengths shorter than these 190 - 270 km (if the phase fronts are aligned with the line of sight; the RO technique might, however, be able to detect GW with shorter horizontal wavelengths than is the case if the phase fronts are perpendicular to the line of sight, see Kursinski et al. (1997) and de la Torre and Alexander (2005) for details). Hence, it is clear that RO-observations are only sensitive to GW with rather large horizontal wavelengths whereas lidar observations may also detect much smaller scale gravity waves. **Note that there is also a (moderate) difference in vertical resolution which is 900 m for the lidar temperatures and $\sim$ 1.4 km for the RO-data (Kaifler et al., 2015; Kursinski et al., 1997).** In addition, we also need to realize that the spatial sampling for both data sets is very different: while the E_P-values based on RO-data are typically based on 20 - 40 single (snapshot) temperature profiles that have been obtained in a geographical area of 5° in latitude and 10° in longitude, the lidar data shown here are based on 10-20 nightly means each consisting of several hours of GW observations (see lower panels in Figure 8). While it is difficult to assess the quantitative impact of this very different sampling on the resulting E_P-values, it is conceivable that the large geographical area over which the RO-data are obtained might result in a smearing out of local GW maxima and should hence tend to smaller values compared to local observations.

In all, we conclude from the comparison of time series at the three considered locations that the fit between GPS-RO and IFS and ERA-Interim data is generally very good whereas comparison to local observations indicates that RO-E_P-values are low-biased - which is likely due to different observational filters of both techniques (see, e.g., Alexander et al., 2010; Ern et al., 2004, for a thorough discussion of observational filters of different techniques). Next, we finally compare GPS-RO with IFS and ERA-Interim data on a global basis. For all thirty months between July 2014 and December 2016 we have computed E_P on a grid of 5° in latitude, 10° in longitude, and 1 km in the vertical for the whole considered altitude range of 20 - 40 km. For each altitude, we have then analyzed the relation between the two GPS-RO data sets and the model data sets in terms of correlation coefficients as well as in terms of absolute differences. An initial impression of the statistical relation between E_P-values from RO-dry data and from IFS-data is presented in Figure 9 which shows corresponding scatter plots along with a linear regression to the data as well as histograms of the absolute difference between the two data sets for three selected altitudes. Figure 9 shows a very large correlation of R=0.94 at 22 km, a minimum value of R=0.45 at 28 km, and a slightly larger value of R=0.56 again at 38 km altitude. Furthermore, it is common to all three histograms that IFS-values are biased low with respect to the RO-data. Interestingly, though, the distribution is broadest at the lowest considered altitude with much narrower distributions above. The complete altitude variation of correlations as well as biases is shown in Figure 10 which shows correlation coefficients as well as median differences (along with 10% and 90% percentiles) between ERA-Interim and IFS, between RO-dry data and IFS-data, and last not least between RO-wet data and IFS-data. Figure 10 shows several interesting features. Starting with the correlation coefficients, those are generally large (between 1,0 and 0.5) except for the altitude range between 25 km and 30 km where the correlation of both RO-data products (wet and dry) with model data show a minimum with values as low as 0.4. Above 30 km, however, correlations coefficients increase again. Besides this striking minimum between 25 km and 30 km, the overall envelope of the altitude variation shows larger correlation coefficients between 0.9 and 1.0 below 25 km and values between 0.8 (for the correlation between ERA Interim and IFS) and 0.5 (for the correlations between the RO-dry data and IFS data) at 40 km. Turning to absolute differences (right panel in Figure 10), the median differences between ERA-Interim and IFS-data are very small (less than 1 J/kg) with IFS-values being slightly larger than ERA-Interim values. Concerning the absolute differences between RO- and IFS-data, both RO data products yield systematically larger E_P-values than the IFS where, however, the median difference between the RO-wet data and the IFS-data is significantly smaller than the difference between the more ‘original’ RO-dry data and the IFS-data. Interestingly, both the median difference as well as its variability (indicated by the percentiles) is quite large at 20 km and decreases significantly up to an altitude of 25 km above which both median differences and related variability increases again up to the maximum altitudes considered.

5 Discussion

In order to identify the reason for the reduced correlation between RO and IFS-data between 25 km and 30 km as well as the relatively large bias below ∼23 km, we next consider a comparison of latitude-longitude distributions of E_P-values at selected altitudes based on RO-dry data and IFS-data. Corresponding results for December 2015 and June 2015 are presented in Figures 11 and 12, respectively. We start with a discussion of the relatively low correlation coefficients at altitudes between 25 and 30 km. Inspection of Figures 11 and 12 reveals that the likely reason for this is that apparently the IFS is hardly simulating any gravity wave activity at the altitude levels of lowest correlation whereas the observations do show some weak but clearly detectable GW activity. The reason why the IFS does not simulate any (respectively very weak) GW activity in the considered vertical wavelength range at these altitudes is not clear at this point but is consistent for all months considered in this study and should be further investigated in the future. As for the bias at altitudes below 25 km, the E_P-distributions shown at 20 and 22 km show that the strongest (apparent) GW activity is here observed in a band of ±20° around the equator with significantly larger values seen in RO-data than in IFS-data. This is, however, the region of the tropical tropopause and its related TTIL. Note that it is on purpose that we refer to the tropical tropopause inversion layer as TTIL instead of the more commonly known TIL, since the latter term has usually only been used for the mid latitude TIL and not the tropical one that we are dealing with here (Birner et al., 2002, 2006; Pilch Kedzierski et al., 2016). That this is indeed the case for the here considered data set is demonstrated in Figure 13 which shows zonal mean N^2-values based on RO and IFS data, respectively. **Note that the N^2-values in Figure 13 were computed from monthly mean zonal mean temperatures which must not be confused with the N^2-values used in our E_P-calculation which is based on T_0-profiles. Remember that T_0-profiles result from filtering individual temperature profiles with a 5th-order Butterworth-filter with cutoff wavelength at 15 km such that T_0-profiles only contain spatial scales larger than 15 km, and hence do not contain information on the TTIL.**

Figure 13 clearly shows that it is indeed the latitude and altitude range of the TTIL which coincides with corresponding regions of large E_P-values in the considered data sets. In addition, Figure 13 also shows that the TTIL is more pronounced in the RO-data than in the ERA-Interim data. Hence, it is tempting to speculate that the large E_P-values seen in the tropics and the corresponding large differences between the RO-data and the IFS-data is because our algorithm to derive E_P-values from temperature profiles by means of separating background temperatures from gravity wave induced disturbances fails in this altitude and latitude region. In order to test this idea further, we present zonal mean E_P-values as a function of latitude and altitude between 20 and 40 km altitude based on both RO-dry-data and IFS-data in Figure 14. This Figure clearly shows the region of large E_P-values between 20 and 25 km altitude and at latitudes between -20° to +20°. It also shows that RO-values in this region are significantly larger than in the IFS-data set. In order to test whether these are indeed real indications of gravity wave activity or rather artifacts due to the TTIL we have next applied our algorithm to derive E_P-values to monthly mean zonal mean temperature profiles. For those, it can safely be assumed that they do not contain any remaining gravity wave signatures (since many profiles have been averaged) such that any significant non-zero E_P-values must be artifacts due to shortcomings of the algorithm. The result of this exercise is shown in the middle panels of Figure 14. Quite obviously this analysis yields regions of very large apparent E_P-values in regions of the TTIL. Compared to the panels in the upper row of the figure, it is also clear that these artifacts actually dominate the E_P-values in the TTIL-region. In addition, we note that additional artifacts are observed at higher altitudes and also in other latitude and altitude regions. These may be caused by tropical Kelvin waves or other planetary scale features such as inertial instability (e.g., Ern et al., 2008; Smith and Riese, 1999). However, for these, their absolute values are significantly less than in the data sets in the upper row such that the contribution of these artifacts to the overall E_P-values is not significant. This is also clearly seen in the lowermost panels of Figure 14 which show the difference of the full E_P-distribution (in the top row) and the contributions from the monthly mean zonal mean profiles (in the middle). In these ‘corrected’ E_P-distributions, the maximum values in the tropical TIL-region have basically disappeared whereas there is hardly any change visible at other altitude and latitude regions. Coming back to the right panel of Figure 10 we hence conclude that the relatively large differences seen below 25km do not reflect real differences in terms of gravity wave activity in RO data and model data. Rather the differences are caused by differences in the representation of the TTIL and the difficulty to properly derive E_P-values in its environment from vertical profiles alone.

We finally attempt to determine the quality of the corrected E_P-values in Figure 14 by comparing them to E_P-values using a horizontal background determination method. Horizontal estimation of T_0 was previously found to be superior to a vertical background determination by Khaykin (2016) and Schmidt et al. (2016). While the sampling statistics of the METOP RO-data on a daily basis (i.e., only 1100 profiles distributed over the whole globe) is too poor to allow us to apply a horizontal background determination to them we may easily perform a corresponding analysis of the high resolution IFS-data. For this purpose the spectral model output of the IFS for December 2015 has been reconstructed at T42, i.e., at a horizontal grid spacing of 500 km. These fields have then been used as background temperatures $T_0(z, \lambda, \phi)$, where λ is latitude and ϕ is longitude, in order to compute monthly mean zonal mean distributions of E_P. Such monthly mean zonal mean E_P-distributions for December 2015 are presented in Figure 15. In the same figure we also show corresponding fields of the vertical kinetic energy, $VE = \frac{1}{2}w^2$ (Geller and Gong, 2010). Note that VE is a

good indicator for gravity waves in the stratosphere since vertical velocities due to other air motions are significantly smaller. While VE-values are significantly smaller than E_P-values (by about a factor of 1000 in the IFS-model) it is still instructive to compare the spatial morphology of the corresponding fields. This comparison clearly reveals that the proposed correction of E_P-distributions derived using a vertical background determination (see Figure 14 and related text) improves the comparison between EP and VE but that it cannot eliminate all features that are apparently not due to gravity waves. Closer inspection of the data sets reveals that this is partly because some of the non-gravity wave structures (mainly the TTIL) are not zonally homogeneous such that correcting for them using zonal mean fields cannot eliminate the non-gravity wave structures completely. We hence conclude that this correction may be recommended for application to data sets that can only be analyzed using a vertical background determination method such as for the METOP data with relatively scarce sampling statistics. However, even after this correction, regions within ±30° latitude around the equator need to be considered with care due to additional potential contamination of E_P by Kelvin waves or other planetary scale features. In any case, if the sampling statistics allows, our analysis clearly shows that in general a horizontal background determination is advantageous in that it better avoids contributions to E_P that are not caused by gravity waves.

6 Summary and Conclusions

In this manuscript we compared operational METOP GPS-RO temperatures and derived gravity wave potential energy densities with corresponding ECMWF operational analysis and ERA Interim reanalysis data sets. This was done to answer two questions, namely whether the sampling and data quality of the operational RO data set is sufficient to properly characterize the global gravity wave activity (measured in terms of E_P) on a monthly basis. Furthermore, we investigated whether the METOP-observations are consistent with the ECMWF model-fields such that the latter can be used for the interpretation of observational results.

For this purpose, we analyzed a total of 30 month of RO data for the period from July 2014 to December 2016. We calculated monthly mean temperatures and E_P-values on a grid of 5° in latitude, 10° in longitude, and at a vertical resolution of 1 km for altitudes between 20 and 40 km. This was done for two RO data sets, namely for so-called ‘dry’ and ‘wet’ data both provided by EUMETSAT’s ROM SAF. Dry temperatures are directly derived from refractivity profiles which in turn are estimated from bending angle observations with the GPS RO-technique. In contrast, wet temperatures are the result of a one dimensional variational retrieval that uses additional a priori information on atmospheric humidity and temperature from ECMWF model fields. Subsequently both temperatures and E_P-values from RO-observations and from ECMWF analysis and reanalysis model fields were compared rigorously.

The comparison of temperatures showed very low systematic differences between RO dry temperatures and ECMWF model fields between 20 and 30 km (i.e., median temperature differences between -0.2 and +0.3 K) which then increased with height to yield median differences of +1.0 K at 34 km and +2.2 K at the maximum considered altitude of 40 km. Compared to this, median differences between RO-wet temperatures and ECMWF-model data were below 0.16 K for all considered altitudes, which is as expected since ECMWF model data were used to constrain the RO data retrieval.

We then introduced a method to derive E_P from temperature profiles by applying a fifth order Butterworth filter with cutoff wavelength of 15 km to both RO and model data. An initial comparison of E_P-time series in selected altitude ranges and at three selected locations in Sodankylä, Northern Scandinavia, in the German Bavarian Forest, and at Lauder, New Zealand, yielded overall very good agreement: below 35 km, this agreement was both very good in terms of seasonal variation and in terms of absolute E_P-values. A striking result, however, was that for Northern Scandinavia - which is known as a region of strong orographic wave activity - the horizontally coarser resolved ERA-Interim data underestimated a large winter peak of E_P that was present in both the RO data as well as in the higher resolution IFS-data. At altitudes above 35 km, however, both models did follow the observed seasonal variation of E_P qualitatively but underestimated the observed values by about a factor of two. This is likely caused by the damping of small scale model structures by the model's sponge layer. **Also, it is well known that noise in RO-data picks up substantially above 35 km such that several previous studies have recommended to restrict the useful range of RO-data for GW analysis to below 35 km (e.g., Schmidt et al., 2008). This previous recommendation is clearly supported by our analysis.**

The same E_P time series from RO-observations were then also compared to local Rayleigh lidar observations. This comparison showed a qualitatively similar seasonal E_P variation with both experimental techniques but it also revealed that the RO-technique underestimates the locally observed values by about a factor of two. This low bias is likely caused by the very different observational filter of RO- and lidar observations where in particular the long line of sight of RO-observations that are carried out in limb geometry severely hampers the detection of waves with horizontal wavelengths smaller than 190 - 270 km while the lidar observations are also sensitive to much smaller horizontal wavelengths.

Finally we compared the full 30 month data set of RO and model E_P-fields. The corresponding statistical analysis shows large correlation coefficients (0.4 - 1.0) between all considered data sets (RO dry, RO wet, ERA Interim, and IFS) for all altitudes between 20 and 40 km. A minimum correlation (of still 0.4) was found at altitudes around 28 km, where the ECMWF-anaylsis and reanalysis fields do not seem to capture the GW-activity that is observed in the RO data. The reason for this discrepancy could not be identified and should be investigated in a future study. Concerning absolute differences between observed and modelled E_P-values, the median difference was relatively small at all altitudes with an exceptional feature between 20 and 25 km where both the median difference between RO- and model-data increased and where also the corresponding variability was found to be very large. The reason for this was identified as an artifact in the E_P-algorithm: this erroneously interprets the pronounced climatological feature of the TTIL at latitudes between ±20° and altitudes between 20 and 25 km as gravity wave activity hence yielding a) very large E_P-values in this area and b) also large differences between model and observations because the RO data show a much more pronounced TTIL than IFS and ERA-Interim. Based on that finding we also suggested a correction for this effect based on an estimate of this 'artificial' E_P using monthly mean zonal mean temperature profiles which do reveal a very pronounced TTIL but which should not contain any remaining GW-signatures due to strong averaging. **In addition, this technique to derive and correct E_P based on vertical profiles was compared to an alternative method applying a horizontal background temperature determination method to IFS-data. We find that the**

above introduced correction may be recommended for application to data sets that can only be analyzed using a vertical background determination method such as the METOP data with relatively scarce sampling statistics. However, if the sampling statistics allows, our analysis also shows that in general a horizontal background determination is advantageous in that it better avoids contributions to E_P that are not caused by gravity waves like the TTIL and potentially also Kelvin waves and other planetary scale features with short vertical wavelengths (i.e., less than 15 km).

In summary, our analysis shows good quantitative agreement between monthly mean RO-dry and ERA-Interim and IFS-data in the altitude range between 20 - 40 km altitude. Hence, both research questions posed at the beginning of this study can be answered positively: for one, this good agreement shows that METOP RO-dry data are a suitable data base to study monthly mean global gravity wave activity in the altitude range between 20 and 40 km (with the caveat that the tropical latitudes need to be considered with particular care). In addition, the good agrement between RO-dry and ECMWF data also implies that the combination of both appears to be a versatile combined data set for the study of processes determining the GW climatology. Future questions to be considered are, for example, in how far the strong stratospheric jet streams influence the observed GW morphology in the stratosphere. While model results of Dunkerton (1984) and more recently also Sato et al. (2009) and Sato et al. (2012) have long suggested that the waves should be refracted into the jet streams, observational evidence for this process based on global data is still scarce. This and other research questions will be investigated in future studies.

7 Data availability

RO data are availabe from www.romsaf.org. Details regarding access to ECMWF-data can be found under www.ecmwf.int.

Author contributions. MR devised the study, wrote the data analysis code, performed the analysis, and drafted the 1st manuscript version. AD provided the ECMWF data and helped with its handling and analysis. BK provided the lidar data and contributed to its analysis. All authors contributed to writing the text.

Competing interests. The authors declare that they have no competing interests.

Acknowledgements. GPS RO data provision by EUMETSAT's ROM SAF is greatly appreciated. Access to the ECMWF data was possible through the special project "HALO Mission Support System". Part of this research was conducted within the scope of the German research initiative ‘Role of the Middle Atmosphere in Climate (ROMIC)’ under grant 01LG1206A provided by the German Ministry for Education and Research. Partial funding was also provided by the German Science Foundation (DFG) via the research unit MS-GWaves (GW-TP/DO 1020/9-1, PACOG/RA 1400/6-1). MR would also like to thank J. Wickert for pointing his attention to METOP RO data and W. Beer for his efficient handling of file downloads from the ROM SAF. MR also thanks P. Preusse and M. Ern for valuable comments on early results achieved within this study and Natalie Kaifler for critically reading and commenting on the text.

References

Alexander, M. J., Geller, M., McLandress, C., Polavarapu, S., Preusse, P., Sassi, F., Sato, K., Eckermann, S. D., Ern, M., Hertzog, A., Kawatani, Y. A., Pulido, M., Shaw, T., Sigmond, M., Vincent, R., and Watanabe, S.: Recent developments in gravity-wave effects in climate models and the global distribution of gravity-wave momentum flux from observations and models, Quart. J. R. Met. Soc., 136, 1103–1124, doi:10.1002/qj.637, 2010.

Birner, T., Dörnbrack, A., and Schumann, U.: How sharp is the tropopause at midlatitudes?, Geophys. Res. Lett., 29(14), 1700, doi:10.1029/2002GL015142, 2002.

Birner, T., Sankey, D., and Shepherd, T. G.: The tropopause inversion layer in models and analyses, Geophys. Res. Lett., 33, L14 804, doi:10.1029/2006GL026549, 2006.

Bölöni, G., Ribstein, B., Muraschko, J., Sgoff, C., Wei, J., and Achatz, U.: The interaction between atmospheric gravity waves and large-scale flows: an efficient description beyond the non-acceleration paradigm, J. Atmos. Sci., 73, 4833–4852, 2016.

de la Torre, A. and Alexander, P.: Gravity waves above Andes detected from GPS radio occultation temperature profiles: Mountain forcing?, Geophys. Res. Lett., 32, L17 815, doi:10.1029/2005GL022959, 2005.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., and Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, Quarterly Journal of the Royal Meteorological Society, 137, 553–597, doi:10.1002/qj.828, http://dx.doi.org/10.1002/qj.828, 2011.

Dunkerton, T.: Inertia-gravity waves in the stratosphere, J. Atmos. Sci., 41, 3396–3404, 1984.

Eckermann, S. D. and Preusse, P.: Global measurements of stratospheric mountain waves from space, Geophys. Res. Lett., 286, 1534– 1537, 1999.

Ehard, B.: Horizontal and vertical propagation of mountain waves from New Zealand into the polar night jet, Ph.D. thesis, Ludwig Maximilians University Munich, 2017.

Ehard, B., Kaifler, B., Kaifler, N., and Rapp, M.: 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, http://www.atmos-meas-tech.net/8/4645/2015/, 2015.

Ern, M., Preusse, P., Alexander, M. J., and Warner, C. D.: Absolute values of gravity wave momentum flux derived from satellite data, J. Geophys. Res., 109, D20 103, doi:10.1029/2004JD004752, 2004.

Ern, M., Preusse, P., Krebsbach, M., Mlynczak, M. G., and Russell III, J. M.: Equatorial wave analysis from SABER and ECMWF temperatures, Atmospheric Chemistry and Physics, 8, 845–869, doi:10.5194/acp-8-845-2008, https://www.atmos-chem-phys.net/8/845/2008/, 2008.

Ern, M., Preusse, P., Gille, J. C., Hepplewhite, C. L., Mlynczak, M. G., III, J. M. R., and Riese, M.: Implications for atmospheric dynamics derived from global observations of gravity wave momentum flux in stratosphere and mesosphere, J. Geophys. Res., 116, D19 107, doi:10.1029/2011JD015821, 2011.

Ern, M., Hoffmann, L., and Preusse, P.: Directional gravity wave momentum fluxes in the stratopsphere from high resolution AIRS temperature data, Geophys. Res. Lett., 43, doi:10.1002/2016GL072007, 2017.

Fetzer, E. J. and Gille, J. C.: Gravity wave variance in LIMS temperatures. Part I: Variability and comparison with background winds, J. Atmos. Sci., 51, 2461–2483, 1994.

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

Fritts, D. C., Smith, R. B., Taylor, M. J., Doyle, J. D., Eckermann, S. D., Dörnbrack, A., Rapp, M., Williams, B. P., Pautet, P.-D., Bossert, K., Criddle, N. R., Reynolds, C. A., Reinecke, P. A., Uddstrom, M., Revell, M. J., Turner, R., Kaifler, B., Wagner, J. S., Mixa, T., Kruse, C. G., Nugent, A. D., Watson, C. D., Gisinger, S., Smith, S. M., Lieberman, R. S., Laughman, B., Moore, J. J., Brown, W. O., Haggerty, J. A., Rockwell, A., Stossmeister, G. J., Williams, S. F., Hernandez, G., Murphy, D. J., Klekociuk, A. R., Reid, I. M., and Ma, J.: The Deep Gravity Wave Experiment (DEEPWAVE): An airborne and ground-based exploration of gravity wave propagation and effects from their sources throughout the lower and middle atmosphere, Bull. Am. Meteorol. Soc., 97, 425–453, doi:10.1175/BAMS-D-14-00269.1, 2016.

Fröhlich, K., Schmidt, T., Ern, M., Preusse, P., de la Torre, A., Wickert, J., and Jacobi, C.: The global distribution of gravity wave energy in the lower stratosphere derived from GPS data and gravity wave modelling: attempts and challenges, J. Atmos. Sol. Terr. Phys., 69, 2238–2248, doi:10.1016/j.jastp.2007.07.005, 2007.

Geller, M., Alexander, M. J., Love, P. T., Bacmeister, J., M.Ern, Hertzog, A., Manzini, E., Preusse, P., Sato, K., Scaife, A. A., and Zhou, T.: A Comparison between Gravity Wave Momentum Fluxes in Observations and Climate Models, J. Clim., 26, 6383–6405, doi:10.1175/JCLI-D-12-00545.1, 2013.

Geller, M. A. and Gong, J.: Gravity wave kinetic, potential, and vertical fluctuation energies as indicators of different frequency gravity waves, Journal of Geophysical Research: Atmospheres, 115, n/a–n/a, doi:10.1029/2009JD012266, http://dx.doi.org/10.1029/2009JD012266, d11111, 2010.

Hei, H., Tsuda, T., and Hirooka, T.: Characteristics of atmospheric gravity wave activity in the polar regions revealed by GPS radio occultation data with CHAMP, J. Geophys. Res., 113, D04 107, doi:10.1029/2007JD008938, 2008.

Hindley, N. P., Wright, C. J., Smith, N. D., and Mitchell, N. J.: The southern stratospheric gravity wave hot spot: individual waves and their momentum fluxes measured by COSMIC GPS-RO, Atmospheric Chemistry and Physics, 15, 7797–7818, doi:10.5194/acp-15-7797-2015, http://www.atmos-chem-phys.net/15/7797/2015/, 2015.

Hoffmann, L., Grimsdell, A. W., and Alexander, M. J.: Stratospheric gravity waves at Southern Hemisphere orographic hotspots: 2003–2014 AIRS/Aqua observations, Atmospheric Chemistry and Physics, 16, 9381–9397, doi:10.5194/acp-16-9381-2016, http://www.atmos-chem-phys.net/16/9381/2016/, 2016.

Holton, J. R. and Alexander, M. J.: The role of waves in the transport circulation of the middle atmosphere, Geophys. Monogr., 123, 21–35, 2000.

Jablonowski, C. and Williamson, D. L.: The Pros and Cons of Diffusion, Filters and Fixers in Atmospheric General Circulation Models, pp. 381–493, Springer Berlin Heidelberg, Berlin, Heidelberg, doi:10.1007/978-3-642-11640-7_13, https://doi.org/10.1007/978-3-642-11640-7_13, 2011.

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

Kaifler, N., Kaifler, B., Ehard, B., Gisinger, S., Dörnbrack, A., Rapp, M., Kivi, R., Kozlovsky, A., Lester, M., and Liley, B.: Observational indications of downward-propagating gravity waves in middle atmosphere lidar data, Journal of Atmospheric and Solar-Terrestrial Physics, p. in press, 2017.

Khaykin, S.: Visiting Scientist Report 29; Retrieval of gravity waves from RO measurements: Capacities and limitations, EUMETSAT, 2016.

Khaykin, S. M., Hauchecorne, A., Mzé, N., and Keckhut, P.: Seasonal variation of gravity wave activity at midlatitudes from 7 years of COSMIC GPS and Rayleigh lidar temperature observations, Geophys. Res. Lett., 42, 1251—-1258, doi:10.1002/2014GL062891, 2015.

Kidston, J., Scaife, A., Hardiman, S., Mitchell, D., Butchart, N., Baldwin, M., and Grey, L.: Stratospheric influence on tropospheric jet streams, storm tracks and surface weather, Nature Geosci., 8, 433–440, doi:10.1038/NGEO2424, 2015.

Kursinski, E. R., Hajj, G. A., Schofield, J. T., Linfield, R. P., and Hardy, K. R.: Observing Earth's atmosphere with radio occultation measurements using the global positioning system, J. Geophys. Res., 102, 23,429–23 465, 1997.

Lindzen, R. S.: Turbulence and stress owing to gravity wave and tidal breakdown, J. Geophys. Res., 86, 9707–9714, 1981.

Lübken, F.-J., Austin, J., Langematz, U., and Oberheide, J.: Introduction to special section on Climate and Weather of the Sun Earth System, J. Geophys. Res., 115, D00I19, doi:10.1029/2009JD013784, 2010.

Malardel, S. and Wedi, N. P.: How does subgrid-scale parametrization influence nonlinear spectral energy fluxes in global NWP models?, J. Geophys. Res., 121, 5395—-5410, doi:10.1002/2015JD023970, 2016.

Marquardt, C. and Healy, S. B.: Measurement noise and stratospheric gravity wave characteristics obtained from GPS radio occultation, J. Met. Soc. Jap., 83, 417–428, 2005.

Nappo, C. J.: An Introduction to Atmospheric Gravity Waves, 2nd ed., Academic Press: Cambridge, MA, USA, 2012.

Pilch Kedzierski, R., Matthes, K., and Bumke, K.: The tropical tropopause inversion layer: variability and modulation by equatorial waves, Atmospheric Chemistry and Physics, 16, 11 617–11 633, doi:10.5194/acp-16-11617-2016, https://www.atmos-chem-phys.net/16/11617/2016/, 2016.

Plougonven, R. and Zhang, F.: Internal gravity waves from atmospheric jets and fronts, Rev. Geophys., 52, doi:10.1002/2012RG000419, 2014.

Poli, P., Healy, S. B., and Dee, D. P.: Assimilation of Global Positioning System radio occultation data in the ECMWF ERA–Interim reanalysis, Quarterly Journal of the Royal Meteorological Society, 136, 1972–1990, doi:10.1002/qj.722, http://dx.doi.org/10.1002/qj.722, 2010.

Preusse, P., Eckermann, S. D., Ern, M., Oberheide, J., Picard, R. H., Roble, R. G., Riese, M., III, J. M. R., and Mlynczak, M. G.: Global ray tracing simulations of the SABER gravity wave climatology, J. Geophys. Res., 114, D08 126, doi:10.1029/2008JD011214, 2009.

Randel, W. J. and Wu, F.: Kelvin wave variability near the equatorial tropopause observed in GPS radio occultation measurements, Journal of Geophysical Research: Atmospheres, 110, n/a–n/a, doi:10.1029/2004JD005006, http://dx.doi.org/10.1029/2004JD005006, d03102, 2005.

ROM SAF: Product user manual: NRT and offline level 1 and level 2 products, EUMETSAT, 2014a.

ROM SAF: Algorithm theoretical baseline document: NRT and offline 1d-var products, EUMETSAT, 2014b.

ROM SAF: Validation report: NRT level 2b and 2c 1d-var products, EUMETSAT, 2014c.

Sato, K., Watanabe, S., Kawatani, Y., Tomikawa, Y., Miyazaki, K., and Takahashi, M.: On the origins of mesospheric gravity waves, Geophys. Res. Lett., 36, L19 801, doi:10.1029/2009GL039908, 2009.

Sato, K., Tateno, S., Watanabe, S., and Kawatani, Y.: Gravity wave characteristics in the Southern Hemisphere revealed by a high-resolution middle-atmospheric general circulation model, J. Atmos. Sci., 69, 1378—1396, doi:10.1175/JAS-D-11-0101.1, 2012.

Scherllin-Pirscher, B., Steiner, A. K., Kirchengast, G., Kuo, Y.-H., and Foelsche, U.: Empirical analysis and modeling of errors of atmospheric profiles from GPS radio occultation, Atmospheric Measurement Techniques, 4, 1875–1890, doi:10.5194/amt-4-1875-2011, https://www.atmos-meas-tech.net/4/1875/2011/, 2011.

Schmidt, T., de la Torre, A., and Wickert, J.: Global gravity wave activity in the tropopause region from CHAMP radio occultation data, Geophys. Res. Lett., 35, L16 807, doi:10.1029/2008GL034986, 2008.

Schmidt, T., Alexander, P., and de la Torre, A.: Stratopsheric gravity wave momentum flux from radio occultations, J. Geophys. Res., 121, 4443–4467, doi:10.1002/2015JD024135, 2016.

Shepherd, T.: Atmospheric circulation as a source of uncertainty in climate change projections, Nature Geosci., 7, 703–7080, doi:10.1038/NGEO2253, 2014.

Skamarock, W. C.: Evaluating Mesoscale NWP Models Using Kinetic Energy Spectra, Monthly Weather Review, 132, 3019–3032, doi:10.1175/MWR2830.1, https://doi.org/10.1175/MWR2830.1, 2004.

Smith, A. K. and Riese, M.: Cryogenic Infrared Spectrometers and Telescopes for the Atmosphere (CRISTA) observations of tracer transport by inertially unstable circulations, Journal of Geophysical Research: Atmospheres, 104, 19 171–19 182, doi:10.1029/1999JD900274, http://dx.doi.org/10.1029/1999JD900274, 1999.

Sutherland, B. R.: Internal Gravity Waves, Cambridge University Press: Cambridge, UK, 2010.

Tsuda, T., Nishida, M., Rocken, C., and Ware, R.: A global morphology of gravity wave activity in the stratosphere revealed by the GPS occultation data (GPS/MET), J. Geophys. Res., 105, 7257–7273, 2000.

Šácha, P., Kuchař, A., Jacobi, C., and Pišoft, P.: Enhanced internal gravity wave activity and breaking over the northeastern Pacific–eastern Asian region, Atmos. Chem. Phys., 15, 13 097–13 112, doi:10.5194/acp-15-13097-2015, http://www.atmos-chem-phys.net/15/13097/2015/, 2015.

Wagner, J., Dörnbrack, A., Rapp, M., Gisinger, S., Ehard, B., Bramberger, M., Witschas, B., Chouza, F., Rahm, S., Mallaun, C., Baumgarten, G., and Hoor, P.: Observed versus simulated mountain waves over Scandinavia – improvement of vertical winds, energy and momentum fluxes by enhanced model resolution?, Atmospheric Chemistry and Physics, 17, 4031–4052, doi:10.5194/acp-17-4031-2017, 2017.

Wu, D. L. and Waters, J. W.: Satellite observations of atmospheric variances: A possible indication of gravity waves, Geophys. Res. Lett., 23, 3631–3634, 1996.

Zhang, Y., Xiong, J., Liu, L., and Wan, W.: A global morphology of gravity wave activity in the stratosphere revealed by the 8-year SABER/TIMED data, J. Geophys. Res., 117, D21 101, doi:10.1029/2012JD017676, 2012.

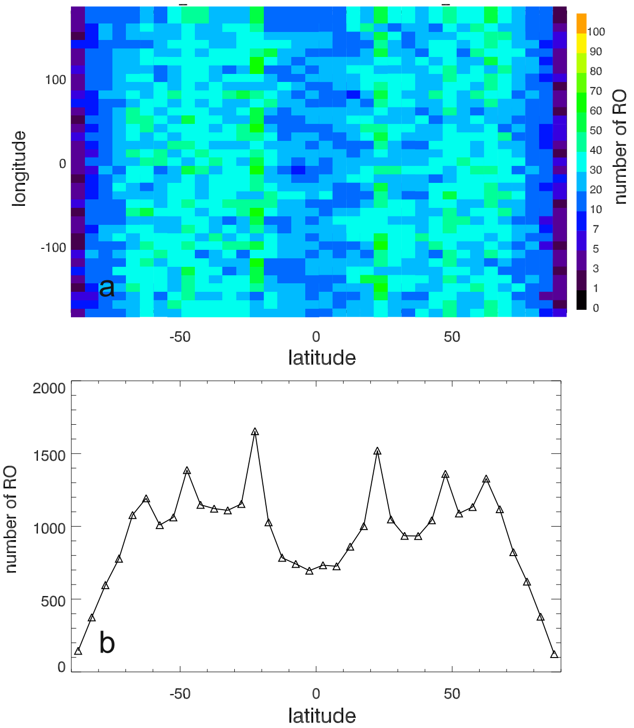


Figure 1. Panel a: Number of Metop A and B radio occultations per 5 degree latitude and 10 degree longitude bin in June 2015. The total number of occultations in this month is about 35.000. Panel b: Number of occultations per 5 degree latitude bin intergrated over all longitudes.

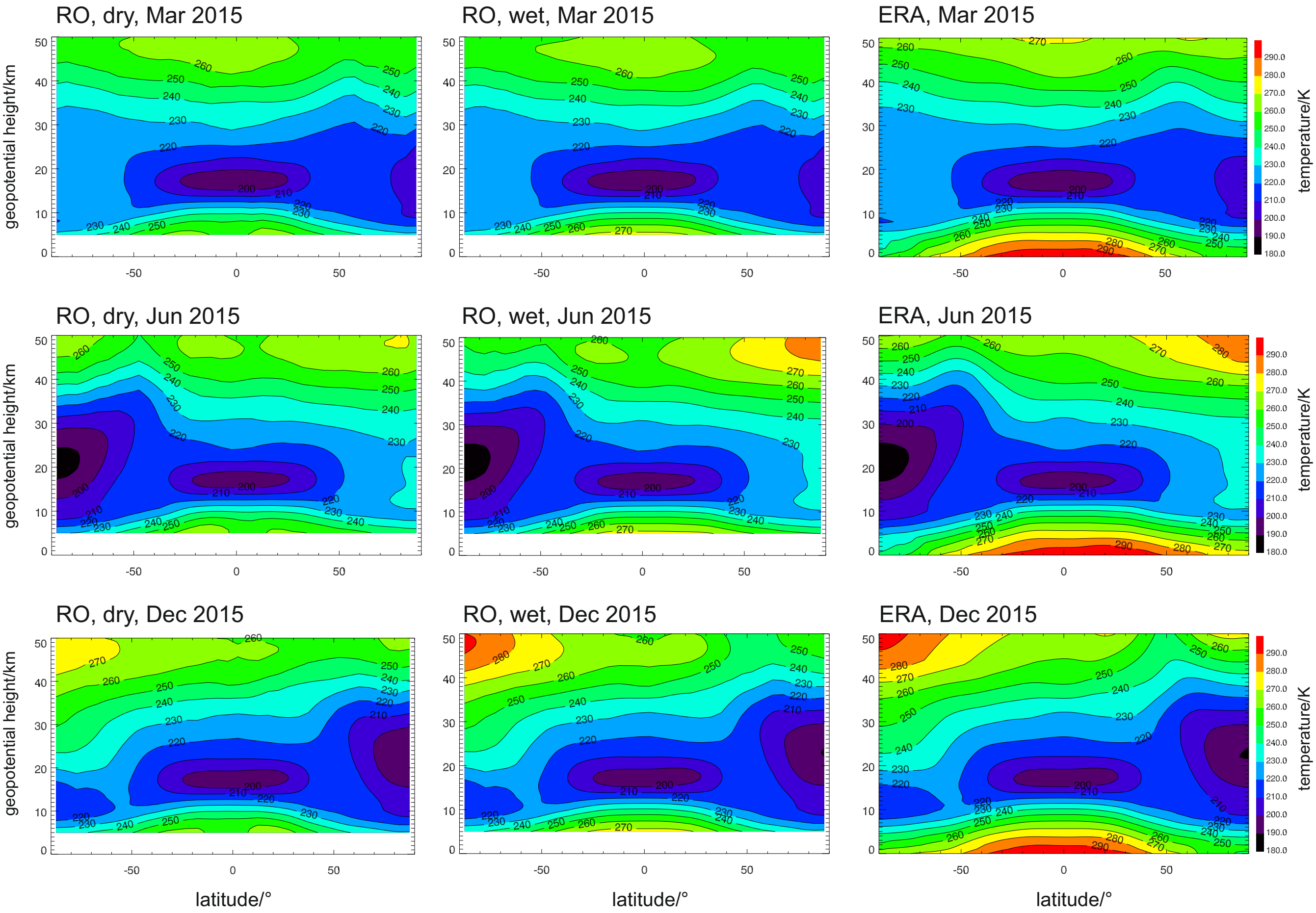


Figure 2. Zonal mean temperatures as a function of latitude and altitude for the months March, June and December 2015 (top to bottom) from Metop A/B radio occultations (left) and from ERA Interim (right).

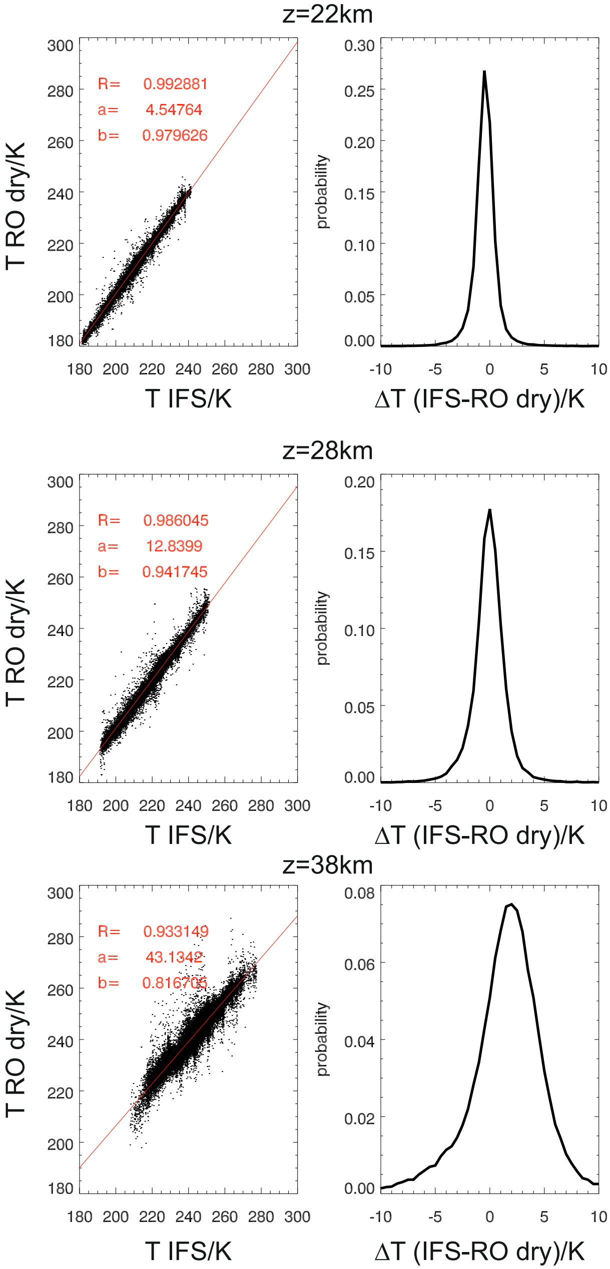


Figure 3. Scatter plots (left) between RO dry temperatures and corresponding IFS-data for 30 months of data between July 2014 and December 2016 for three selected altitudes. The red line shows a linear fit to the data with slope b, y-intercept a, and correlation coefficient R (see insert). Panels on the right show histograms of the corresponding temperature differences between IFS and RO dry data for the same selected altitudes.

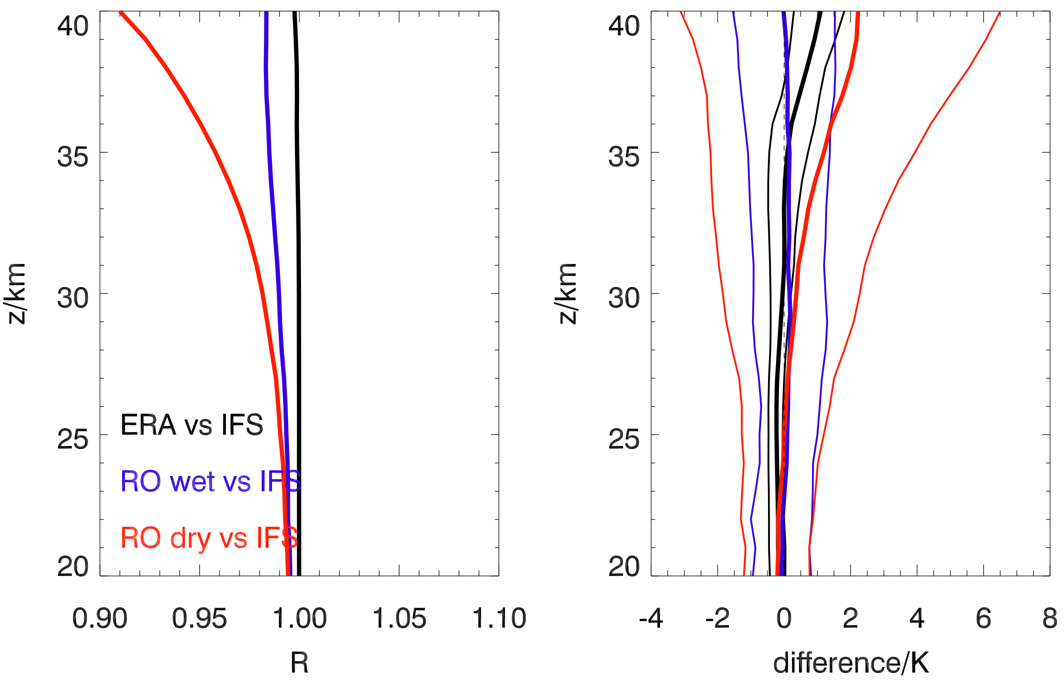


Figure 4. Left: Correlation coefficients as a function of altitude for the correlation between ERA Interim and IFS-data (black line), RO wet- and IFS-data (blue line), and RO-dry and IFS-data (red line). Right: Corresponding median temperature differences (thick lines) along with 10% and 90% percentiles (thin lines) as a function of altitude (same color code as in left panel).

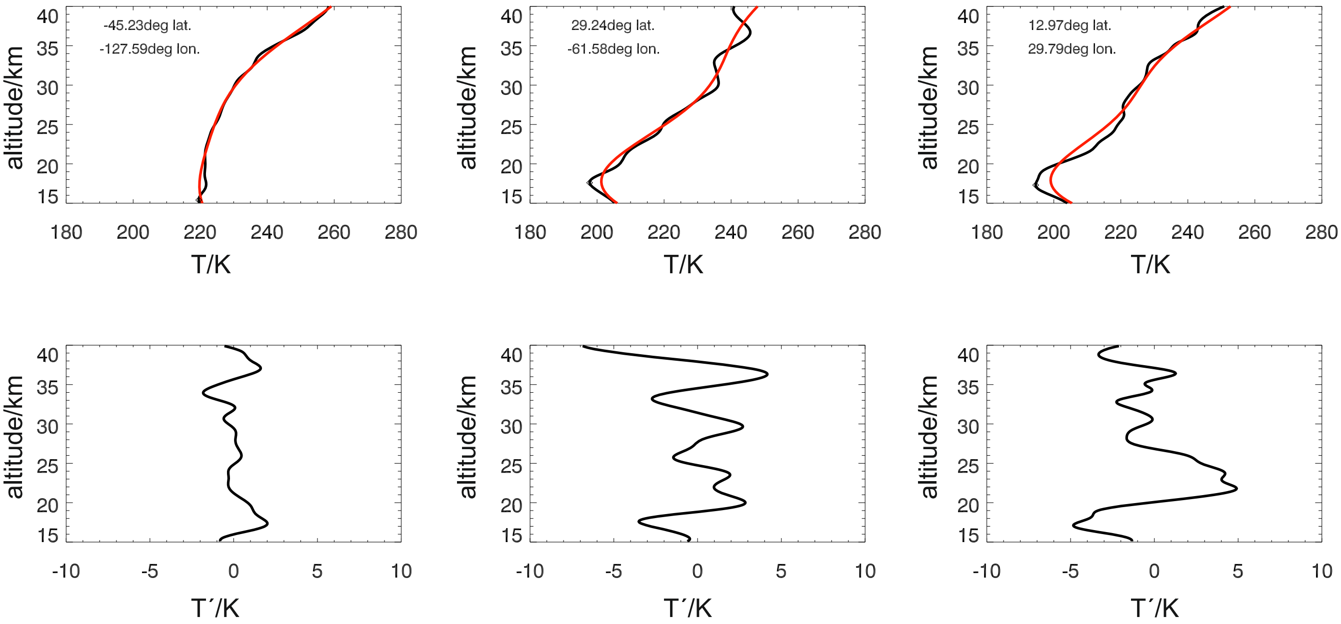


Figure 5. Upper panels: Sample radio occultation temperature profiles from December 2015 (black lines) with background profiles (red lines) as determined with a 5th order Butterworth filter with 15 km cuttoff wavelength following Ehard et al. (2015). Lower panels: corresponding temperature perturbation profiles (= radio occultation profile minus background profile).

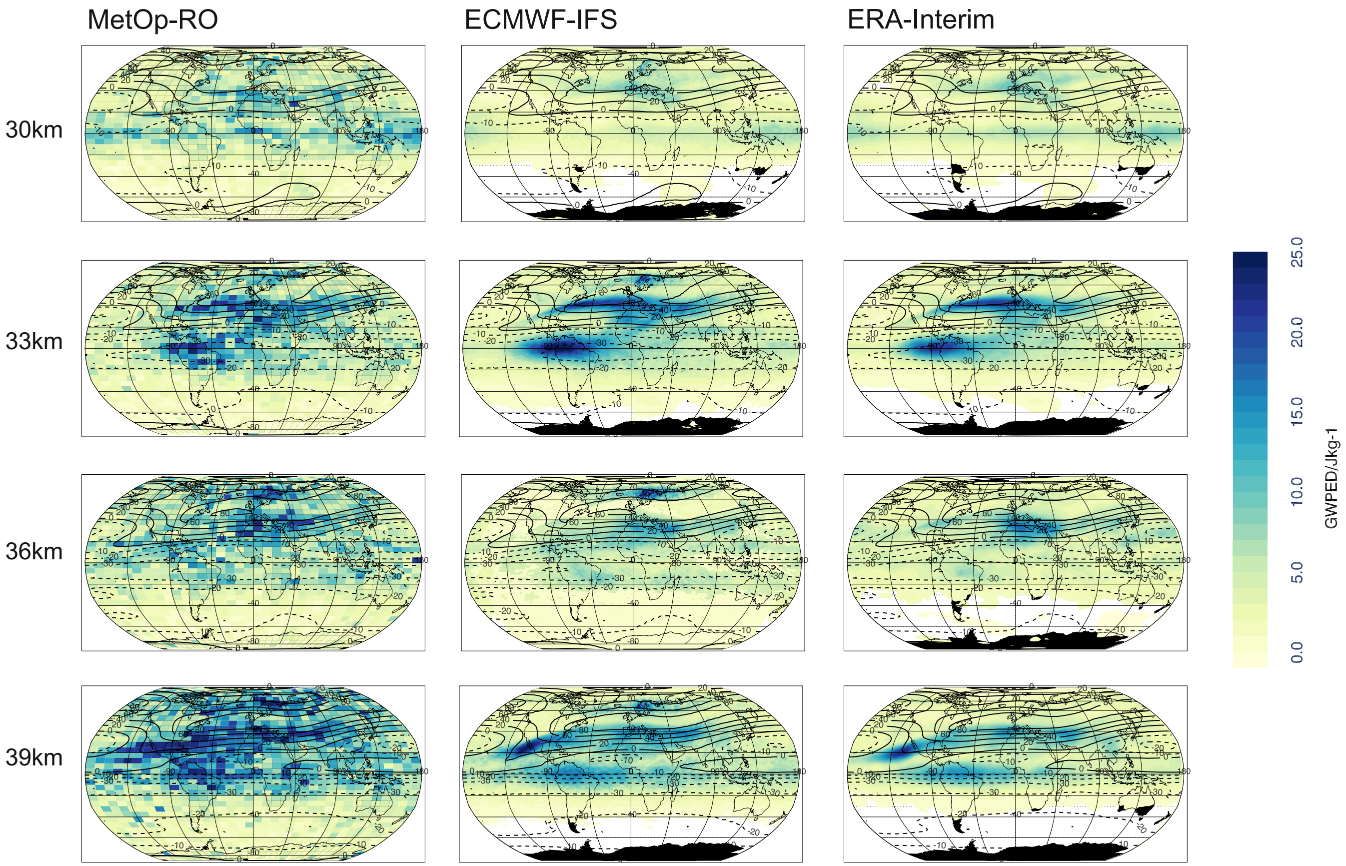
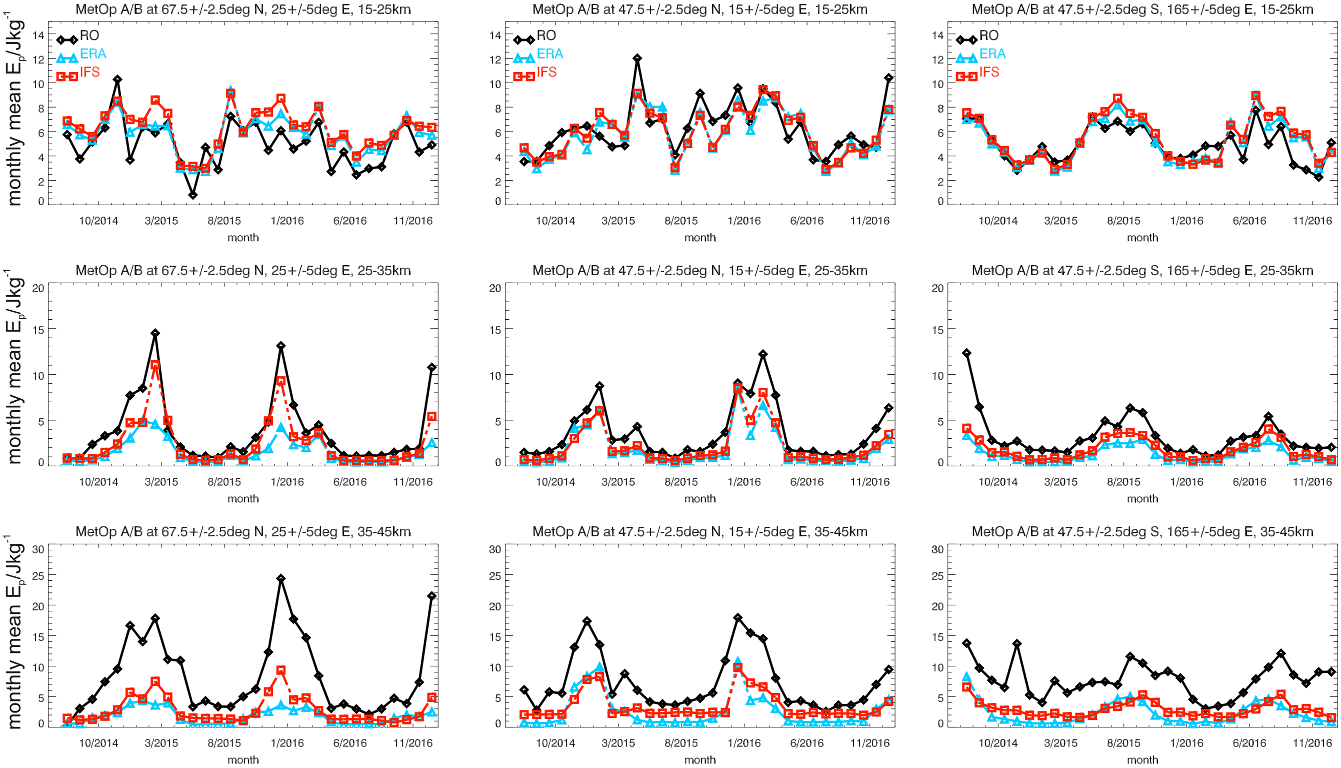


Figure 6. Monthly mean latitude-longitude cross sections of E_p at selected altitudes of 30, 33, 36, and 39 km (from top to bottom) for December 2015. Left panels show METOP-RO dry-data, middle panels show IFS-data, and right panels show ERA-Interim data, respectively. In all panels, black contour lines show zonal wind values from ERA-Interim.



MetOp A/B at 67.5+/-2.5deg N, 25+/-5deg E, 15-25km

MetOp A/B at 47.5+/-2.5deg N, 15+/-5deg E, 15-25km

MetOp A/B at 47.5+/-2.5deg S, 165+/-5deg E, 15-25km

MetOp A/B at 67.5+/-2.5deg N, 25+/-5deg E, 25-35km

MetOp A/B at 47.5+/-2.5deg N, 15+/-5deg E, 25-35km

MetOp A/B at 47.5+/-2.5deg S, 165+/-5deg E, 25-35km

MetOp A/B at 67.5+/-2.5deg N, 25+/-5deg E, 35-45km

MetOp A/B at 47.5+/-2.5deg N, 15+/-5deg E, 35-45km

MetOp A/B at 47.5+/-2.5deg S, 165+/-5deg E, 35-45km

Figure 7. Comparison of time series of monthly mean E_P from RO and model data (ERA-Interim and IFS) for three different locations: Sodankylä (left), Bavarian Forest (middle), and Lauder (right). The color code is explained in the insert. Panels shown in top, middle, and bottom row are for altitude ranges of 15-25 km, 25-35 km, and 35-45 km, respectively.

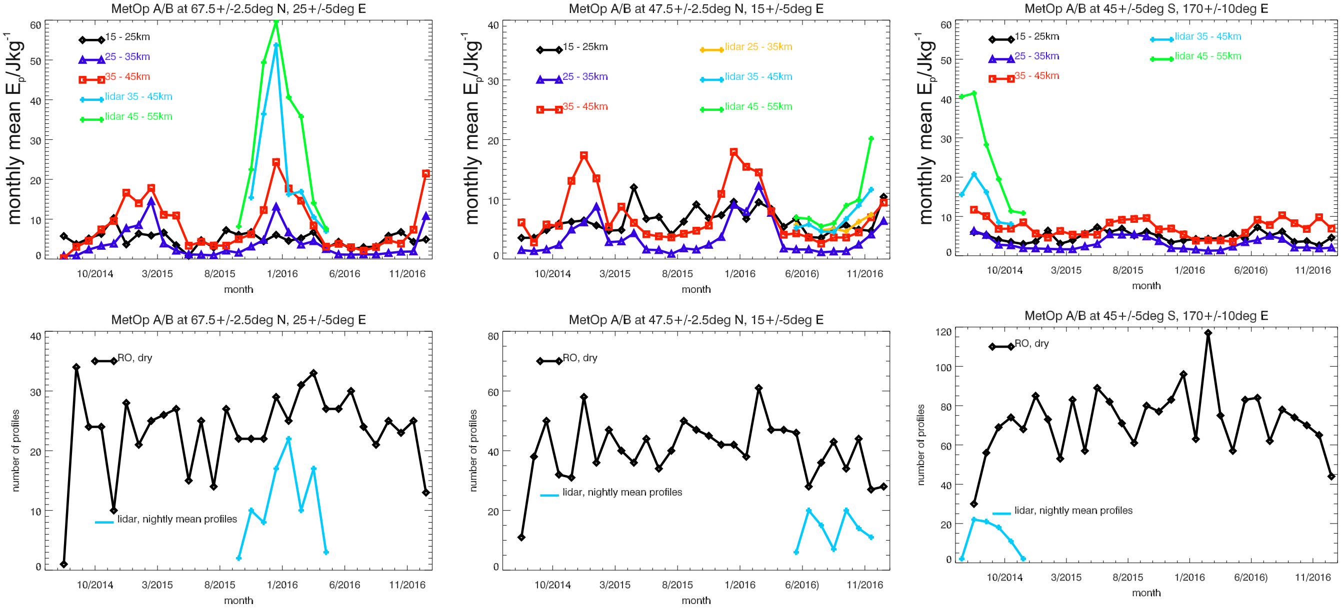


Figure 8. Upper panels: Comparison of time series of monthly mean E_P from METOP-RO data for different altitude ranges (black, blue, and red curves; see insert for color code) with local Rayleigh lidar measurements of E_P for the stations of Sodankylä (left), the Bavarian Forest (middle), and Lauder (right). Lidar E_P are shown as yellow (25-35 km), light blue (35-45 km), or green (45-55 km) lines. Lower panels: Number of RO-profiles (black lines) and nightly mean lidar profiles (light blue lines) entering the monthly mean shown in the panels above.

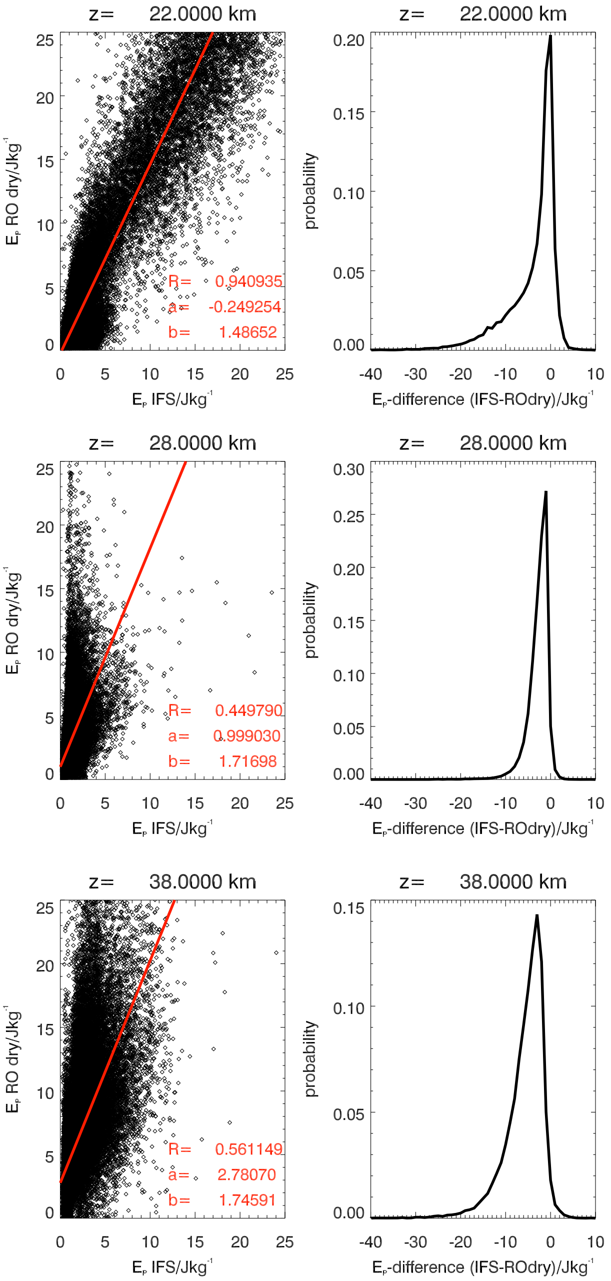


Figure 9. Left: Scatter plots between E_P-values derived from the IFS and RO-dry-data for three different altitudes, i.e., 22 km (top), 28 km (middle), and 38 km (bottom).The red line shows a linear fit to the data with slope b, y-intercept a, and correlation coefficient R (see insert). The panels on the right show corresponding histograms of the difference between the two data sets.

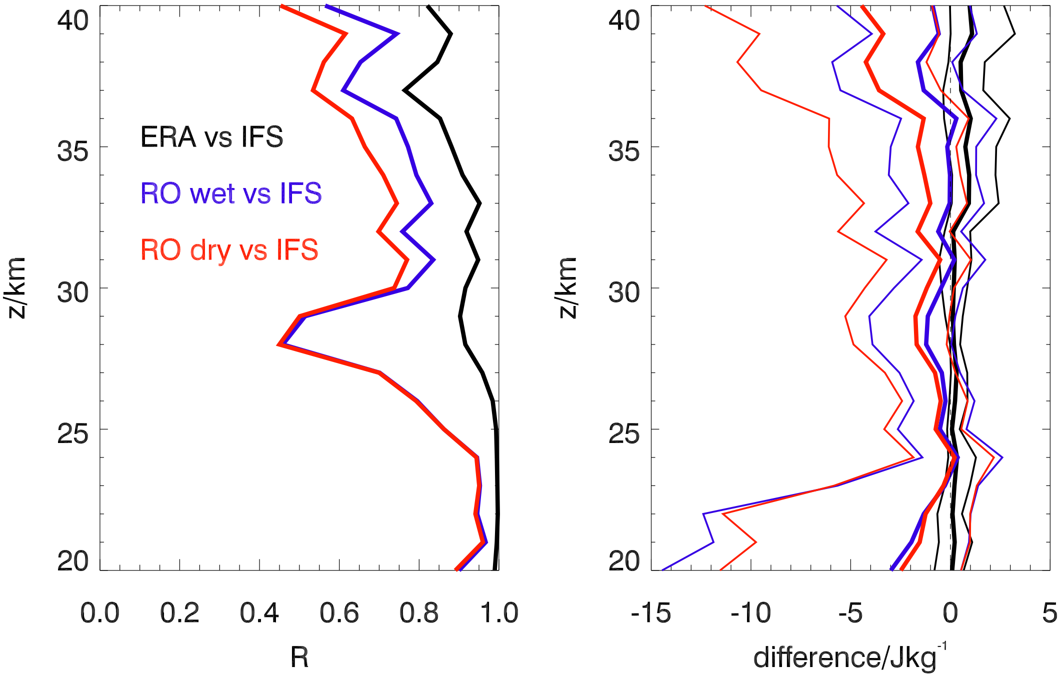


Figure 10. Left: Correlation coefficients as a function of altitude for the correlation between ERA Interim and IFS-data (black line), RO wet- and IFS-data (blue line), and RO-dry and IFS-data (red line). Right: Corresponding median temperature differences (thick lines) along with 10% and 90% percentiles (thin lines) as a function of altitude (same color code as in left panel).

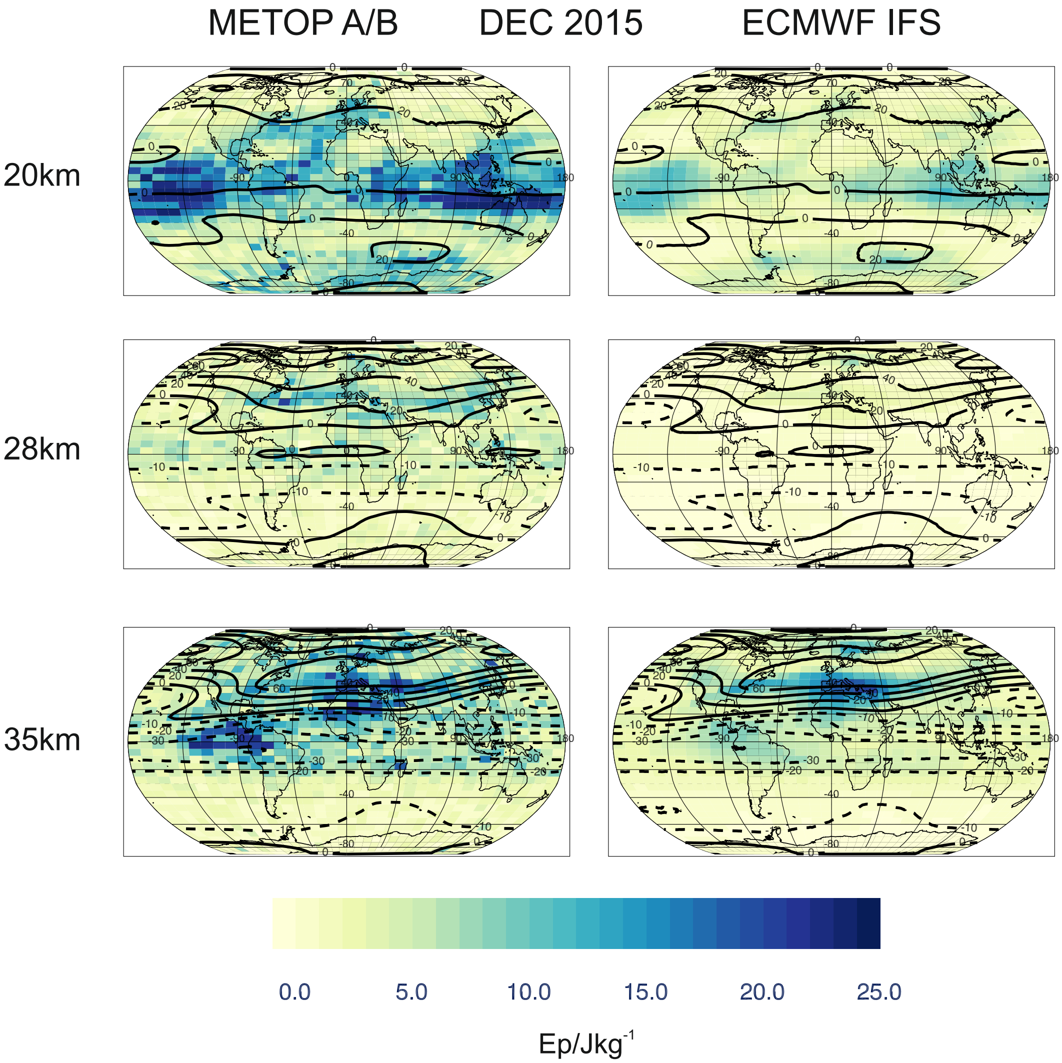


Figure 11. Left: Latitude-longitude distributions of E_P based on GPS-RO-dry-data for December 2015 and altitudes of 20, 28, and 38 km (top to bottom). Right: Same as left panels but based on IFS data. In all panels black contours show zonal wind values from ERA-Interim.

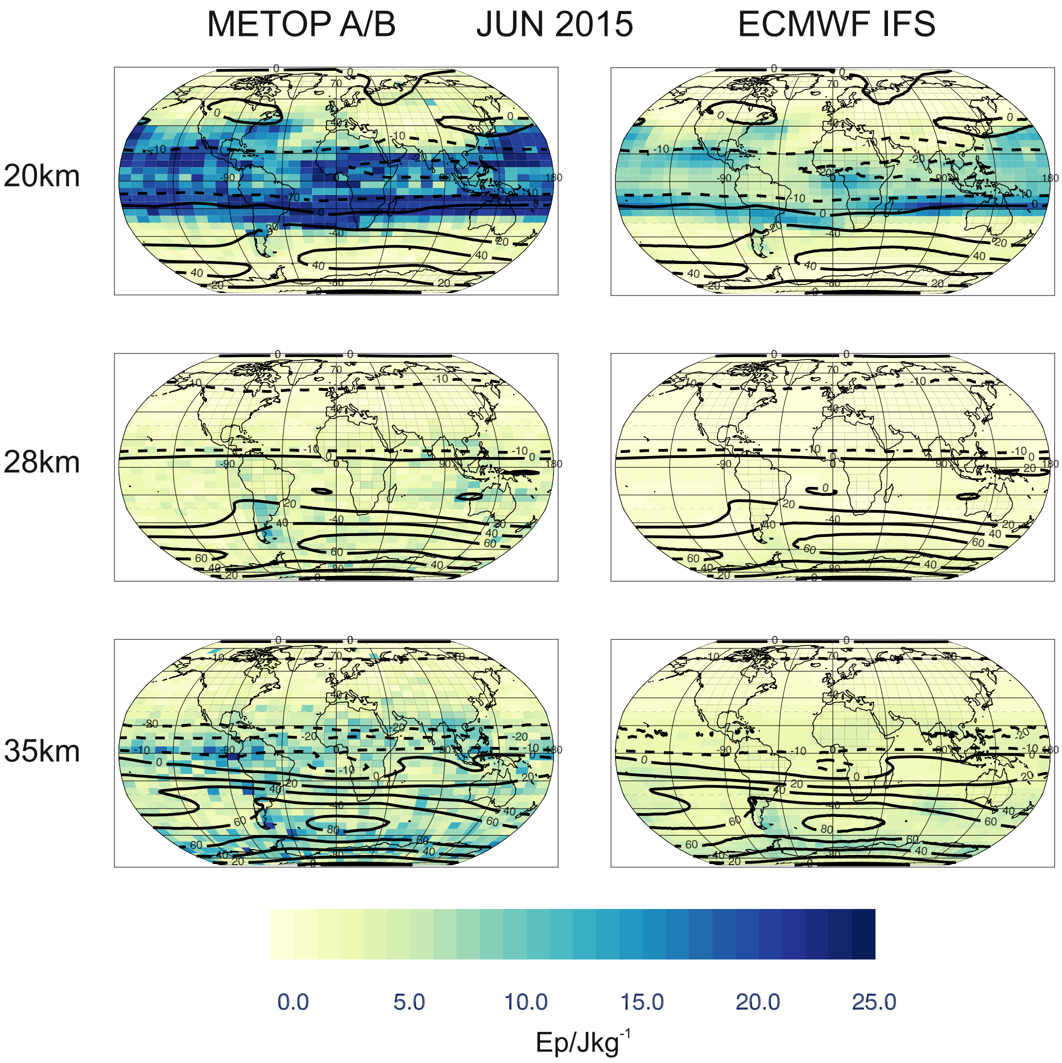


Figure 12. Same as Figure 11 but for June 2015.

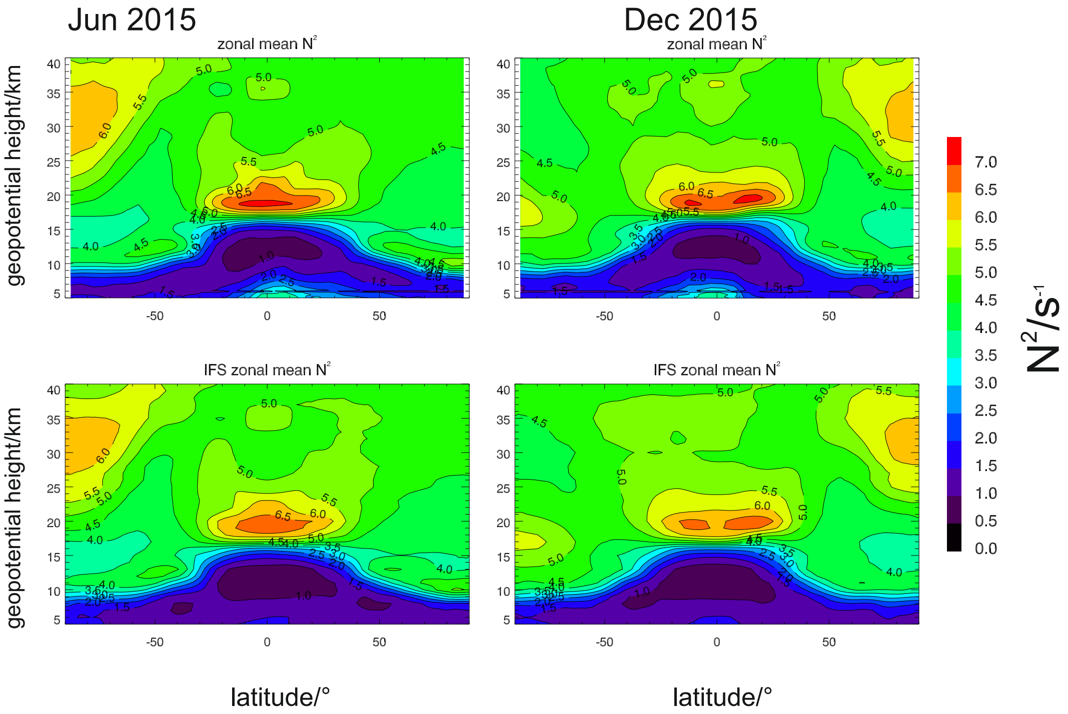


Figure 13. Zonal mean distribution of N^2 as a function of latitude and altitude for the months June 2015 (left) and December 2015 (right) based on GPS-RO-dry-data (upper panels) and IFS data (lower panels), respectively.

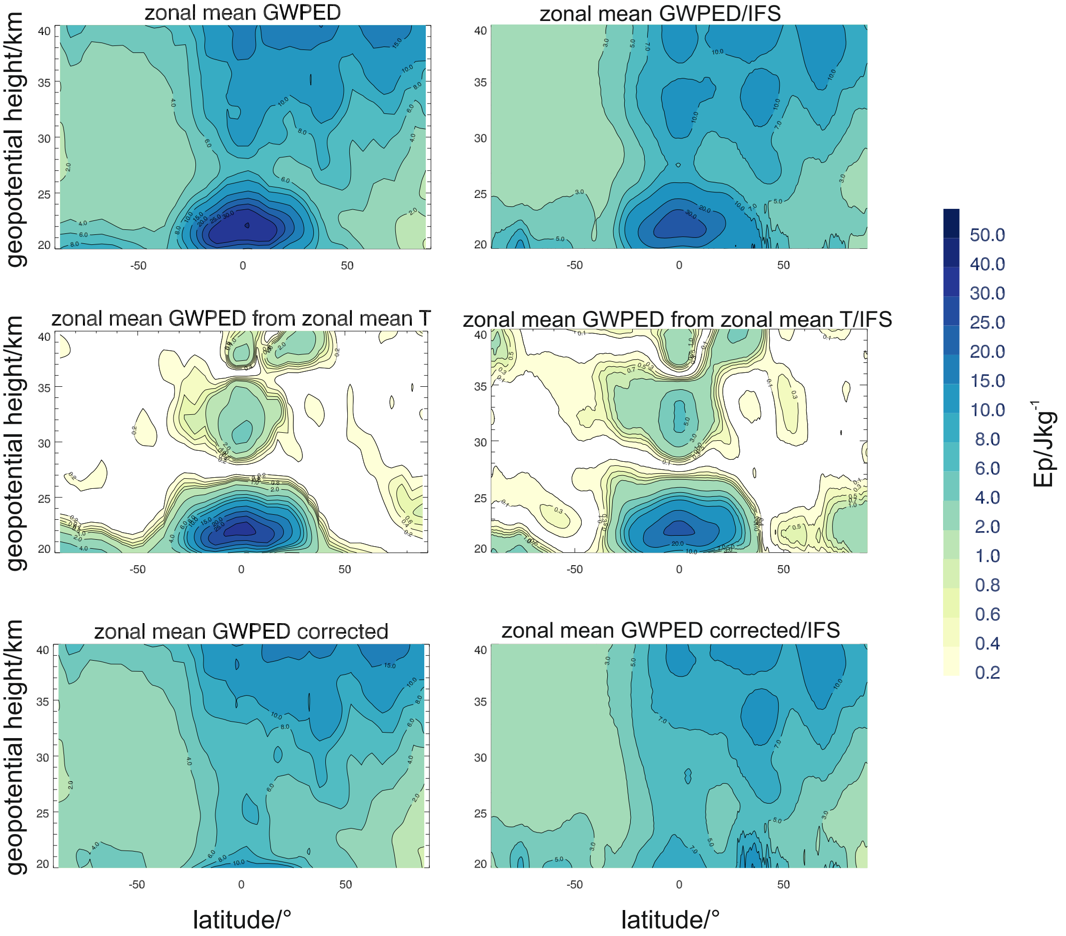


Figure 14. Upper panels: Monthly mean zonal mean distributions of E_P as a function of latitude and altitude for December 2015 based on RO-dry-data (left) and IFS-data (right). Middle panels: Zonal mean apparent E_P-values derived from applying the E_p-algorithm to monthly mean zonal mean temperature profiles. Bottom: Difference between upper panels and middle panels.

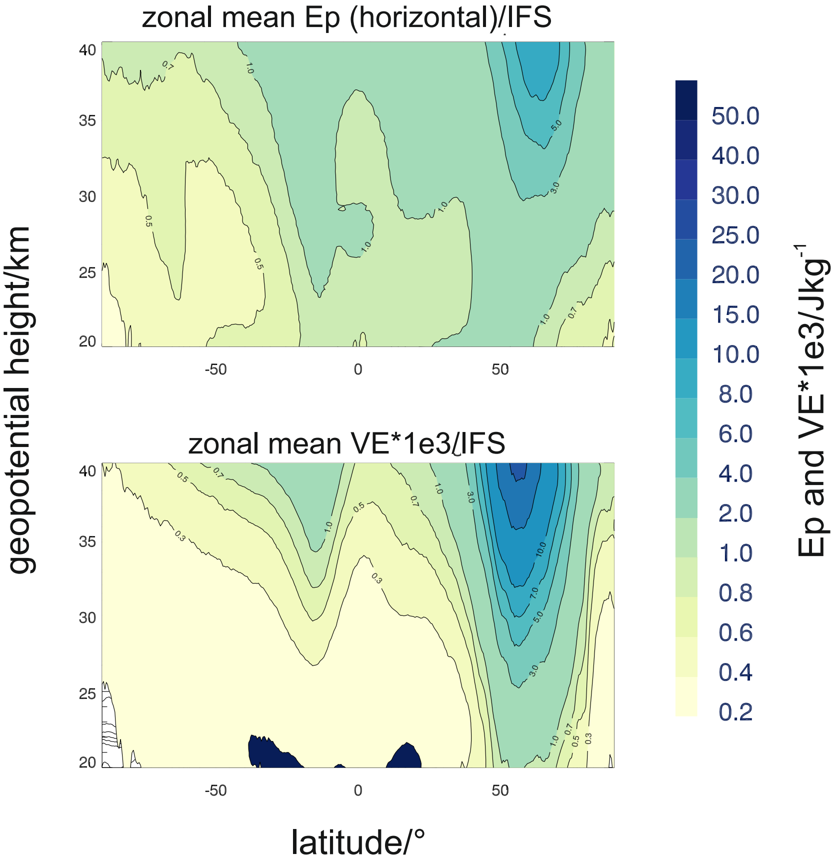


Figure 15. Upper Panel: Monthly mean zonal mean distribution of E_P from IFS data derived after detrending in the horizontal with T42-IFS-fields. Lower Panel: Monthly mean zonal mean distribution of $VE = \frac{1}{2}w^2$.