Using a ~~global~~ network of temperature lidars to identify temperature biases in the upper stratosphere in ECMWF reanalyses

Graeme Marlton[1], Andrew Charlton-Perez[1], Giles Harrison[1], Inna Polichtchouk[2], Alain Hauchecorne[3], Philippe Keckhut[3], Robin Wing[3], Thierry Leblanc[4], and Wolfgang Steinbrecht[5]

[1]Department of Meteorology, University of Reading, Reading, RG6 6LA
[2]European Centre for Medium Range Weather Forecasts, Shinfield Road, Reading, United Kingdom
[3]LATMOS/IPSL, UVSQ Université Paris-Saclay, Sorbonne Univerités, CNRS, Guyancourt, France
[4]JPL-Table Mountain Facility, 24490 Table Mountain Road, Wrightwood, CA., USA
[5]Deutscher Wetterdienst, Albin-Schwaiger-Weg 10, 82383, Hohenpeissenberg, Germany

Correspondence: Graeme Marlton (graeme.marlton@reading.ac.uk)

Abstract. To advance our understanding of the stratosphere, high quality observational datasets of the ~~upper atmosphere~~ stratosphere are needed. It is commonplace that reanalysis ~~is~~ datasets are used to conduct stratospheric studies. However the accuracy of ~~the standard reanalysis~~ these reanalyses at these heights is hard to infer due to a lack of in-situ measurements. Satellite measurements provide one source of temperature information. As some satellite information is already assimilated into reanalyses, the direct comparison of satellite temperatures to the reanalysis is not truly independent. Stratospheric lidars use Rayleigh scattering to measure density in the middle and upper atmosphere, allowing temperature profiles to be derived for altitudes from 30km (where Mie scattering due to stratospheric aerosols becomes negligible) to 80-90km (where the signal-to-noise begins to drop rapidly). The Network for the Detection of Atmospheric Composition Change (NDACC) contains several lidars at different latitudes that have measured atmospheric temperatures since the 1970s, resulting in a long running upper-stratospheric temperature dataset. These temperature datasets are useful for validating reanalysis datasets in the stratosphere, as they are not assimilated into reanalyses. Here we take stratospheric temperature data from lidars in the northern hemisphere ~~for winter months~~ between 1990-2017 and compare them with the European Centre for ~~ECMWF's~~ Medium range Weather Forecasting ERA-interim and ~~ERA-5~~ ERA5 reanalyses. To give confidence in any bias found, temperature data from NASA's EOS Microwave Limb Sounder is also compared to ERA-interim and ~~ERA-5~~ ERA5 at points over the lidar sites. In ERA-interim a cold bias of -3 to -4 K between 10 hPa and 1 hPa is found when compared to both measurement systems. Comparisons with ~~ERA-5~~ ERA5 found a small bias of magnitude 1 K which varies between cold and warm bias with height between 10 hPa and ~~3~~1 hPa, indicating a good thermal representation of the ~~upper atmosphere to 3~~middle atmosphere up to 1 hPa. ~~At heights above this, comparisons with EOS MLS yield a slight warm bias and the temperature lidar yield a cold bias.~~ A further comparison is undertaken looking at the temperature bias by year to see the effects of the assimilation of the Advanced Microwave Sounding Unit-A satellite data and the Constellation Observing System for Meteorology, Ionosphere, and Climate GPS Radio Occulation (COSMIC GPSRO) data on stratospheric temperatures. ~~By comparing periods before and after~~ It if found that ERA5 is sensitive to the introduction of ~~each data source it is clear that COSMIC GPSRO improves the cold bias~~

~~in the 3 hPato 0.5 hPa altitude range~~ COSMIC GPRSO in 2007 with the reduction of the cold bias above 1 hPa. In addition to this, the introduction of AMSU-A data caused variations in the temperature bias between 1-10 hPa between 1997-2008.

1 Introduction

The ~~middle atmosphere~~ stratosphere influences the weather and climate in the troposphere ~~e. g Domeisen (2019); Domeisen et al. (2019a) .~~ (Domeisen, 2019; Domeisen et al., 2019a). Sudden Stratospheric Warmings (SSWs) can cause changes in the tropospheric flow for many weeks ~~e.g. Charlton and Polvani (2007)~~ (Charlton and Polvani, 2007) and the Quasi Biennial Oscillation (QBO), a 28 month switching in equatorial stratospheric winds, also affects the large scale processes in the troposphere ~~e. g. Baldwin et al. (2001) .~~ (Baldwin et al., 2001). Critical to our understanding of how these processes work are good observational datasets of the ~~upper~~ middle atmosphere. Often reanalysis datasets such as the European Centre for Medium range Weather Forecasts' (ECMWF) ReAnalaysis (ERA) ~~datasets~~ (Dee et al., 2011) or the US National ~~Centre~~ Center for Atmospheric Research (NCAR) ~~reanalysis~~ ReAnalysis (Kalnay et al., 1996) amongst many others are used for stratospheric studies on a global scale as shown for example in Fujiwara et al. (2017); Seviour et al. (2012); Lee et al. (2019); Skerlak et al. (2014); Butler et al. (2015).

To create ~~reanalysis~~ reanalyses datasets, a large amount of temperature~~data is~~ , ozone and wind observations are assimilated from satellite and in-situ measurements. In the middle and upper stratosphere the number of temperature observations is somewhat limited. This makes diagnosing bias in a reanalysis dataset more difficult. Radiosondes, small balloon borne instrument packages, which provide in-situ temperature profiles up to heights of about 30 km are launched from thousands of locations daily, giving a wealth of information that is assimilated in the lower and middle ~~atmosphere~~stratosphere. Simmons et al. (2020) undertook a study examining the performance of ~~ERA-5~~ERA5, the ECMWF's most recent reanalysis dataset using radiosonde ~~data. However~~and satellite observations. However, there were height limitations due to the maximum height of available radiosonde data~~,~~ . The technique also had the potential to be biased due to the assimilation of radiosonde observations ~~within~~ into the reanalysis. Rocketsondes (Schmidlin, 1981) can reach heights of 100 km providing high resolution temperature information in the ~~middle and upper atmosphere~~upper stratosphere, mesosphere and thermosphere. Rocketsondes are not operationally assimilated due to the significant installations and costs to launch, which ~~leads~~ has led to sparse temporal and spatial sampling, with the last known campaign occurring in 2004 (Sheng et al., 2015).

There are numerous satellite techniques to retrieve temperature profiles of the ~~upper atmosphere~~stratosphere. Stratospheric sounding units (SSU) (Miller et al., 1980) derive temperature from radiances in CO_2 emissions. Similarly the Sounding of the Atmosphere using Broadband Emission Radiometry (SABER) instrument uses limb emissions from CO_2 to provide temperature observations in the mesosphere and thermosphere (Russell et al., 1999). The Aqua satellite combines data from Atmospheric Infra-red Sounders (AIRS) with data from the Advanced Microwave Sounding Units (AMSU) to provide temperature profiles in the troposphere and stratosphere ~~with 1 km vertical resolution~~ (Susskind et al., 2006). The Microwave

Limb Sounder (MLS) provides temperature data by observing the limb emission of several atmospheric gases and aerosols (Waters et al., 2006). Low Earth orbiters such as those in the Constellation Observing System for Meteorology, Ionosphere and Climate (COSMIC) can derive atmospheric properties such as temperature, pressure and water vapour using GPS Radio Occulation (GPSRO) up to ~~40~~40-50 km (Kuo et al., 2000). As COSMIC is a constellation of satellites it retrieves thousands of randomly sampled temperature profiles daily across the globe. Many ~~such techniques~~ of the above observations are assimilated into ~~reanalysis~~ reanalyses making it hard to make an unbiased comparison.

Another source of ~~middle atmosphere~~ stratospheric temperature measurements is from Rayleigh temperature lidars. These temperature lidars use Rayleigh scattering properties of the atmosphere above the stratospheric aerosol layer (>30 km) to infer density and, by assuming hydrostatic equilibrium, temperature (Hauchecorne and Chanin, 1980). There are several Rayleigh lidars across the globe based at participating sites in the Network for the Detection of Atmospheric Composition Change (NDACC). A small handful have been making temperature profile measurements between 30 km and 90 km for at least three decades. Furthermore the lidar temperature profiles are not assimilated into ~~reanalysis~~ reanalyses making them independent for numerical dataset comparisons. Le Pichon et al. (2015) compared 6 months of Rayleigh temperature lidar data with ECMWF reanalysis data and found good agreement.

In this paper, we ~~use~~ compare temperature data from four ~~lidars and compare it to temperature data from~~ ground based Rayleigh lidars, an independent measurement technology, and use them to infer the stratospheric temperature bias in the ECMWF's ERA-interim and ERA5 ~~reanalysis packages over three decades to identify temperature bias for stratospheric studies~~reanalyses. To add confidence to the identified bias the same comparison will also be undertaken with temperature data from the National Aeronautics and Space Administrations (NASA) Earth Observing System Microwave Limb Sounder (EOS MLS). EOS MLS is one of the few satellite temperature datasets which is not assimilated into ECMWF reanalysis.

The reanalysis packages cover long time periods over which the quantity and types of data assimilated has increased. Poli et al. (2010) showed that inclusion of GPSRO data improved temperature bias in the ~~upper stratosphere and lower troposphere. Hence, further analysis will be undertaken comparing ERA-interim and ERA5 before and after the inclusion of GPSRO data from the COSMIC constellation at the end of 2006.~~ lower stratosphere and upper troposphere. Simmons et al. (2020) also found that the inclusion of AMSU-A data in 2000 caused ~~a increase in cold bias in ERA-5, which we examine here also.~~ an increase in the warm bias in ERA5 at heights above 7 hPa. The four temperature lidars used here span at least 25 years. Hence, further analysis will be undertaken to ascertain how ERA-interim and ERA5's stratospheric temperature bias evolve over the period 1990-2017 with the introduction of both COSMIC GPSRO and AMSU-A data.

2 Dataset description

2.1 Stratospheric Temperature Lidar

Atmospheric lidar remote sensing uses light scattering ~~of the~~ from molecules and particles. A laser pulse is emitted into the atmosphere where it is scattered and absorbed by the molecules and particles. The fraction of light backscattered towards the instrument on the ground is collected by a telescope and sampled as a function of time, which, knowing the speed of light,

translates into altitude. In the absence of particulate matter (typically above 30-35 km), and after several corrections (e.g., non-linearity and range corrections, background noise extraction, molecular extinction and absorption corrections), the lidar signal is proportional to the air density. With the assumption that the atmosphere is an ideal gas in hydrostatic balance, the atmospheric temperature profile is then retrieved by integrating the measured relative density downward from the highest usable data point (Hauchecorne and Chanin, 1980). At the top of the profile (typically, in the mesosphere), ~~an~~ a priori temperature~~and~~ , density or pressure information is needed to initialize the downward integration.

At mesospheric altitudes, empirical models such as the Committee on Space Research's International Reference atmosphere (~~CIRA) -86~~ CIRA-86) (Chandra et al., 1990) or the Naval Research Laboratory's Mass Spectrometer and Incoherent Scattering Radar model (NRL MSISE-00) (Picone et al., 2002) are typically used for the a priori information. Because of tides and gravity waves, mesospheric temperatures can be highly-variable at small spatio-temporal scales (Jenkins et al., 1987), and these models often do not represent well the state of the atmosphere measured by lidar at a given time and location, resulting in a temperature uncertainty of up to 10-20 K at the altitude of initialization. This uncertainty decreases exponentially as the profile is integrated downward, resulting in a typical temperature uncertainty of 1-2 K 15 km below the top of the profile due to the a priori information (Leblanc et al., 2016). In order to avoid misinterpretation of the lidar profile and its influence by the a priori information, the top 10 km of the profiles are often excluded from published datasets (Wing et al., 2018).

Another important source of temperature uncertainty is signal detection noise which ~~is negligible at the~~ has two components. The first is a random component in the form of photon detection, which becomes negligible when averaging, and a systematic component in the form of background noise such as level of sky light levels etc., which can be budgeted and corrections applied for (e.g. Leblanc et al. (2016)). At the bottom of the profile ~~where the~~ the random temperature uncertainty is negligible as the signal-to-noise ratio is high, but can increase to 10 K at the top of the profile where signal-to-noise ratio ~~lowers~~decreases. Because of its random nature, vertical and temporal averaging can reduce detection noise significantly. Background noise correction and signal non-linearity correction are two other sources of uncertainty. Just like a priori and detection noise uncertainty, the ~~background noise correction~~ corrections for background noise uncertainty is negligible at the bottom of the profile and increases as we approach the top of the profile. On the other hand, uncertainty owed to non-linearity correction maximizes at the bottom of the profile (typically less than 2 K), and becomes negligible a few km above.

It should be stressed that each lidar instrument is different and the various contributions to total uncertainty can vary widely depending on the instrument considered (Leblanc et al., 2016). For the comparisons undertaken here, we selected the four longest datasets of the dozen backscatter temperature lidar datasets available at NDACC. These datasets span at least ~~10~~ 25 years and have frequent temperature profiles during that period (see table 1). The instruments are located at the German Weather Service ~~observatory~~ Observatory of Hohenpeissenberg (HOH) ~~(Steinbrecht et al., 2009)~~(Steinbrecht et al., 2009; Wing et al., 2020a) , Germany, the Observatoire de Haute Provence (OHP) ~~(Hauchecorne, 1995)~~(Hauchecorne, 1995; Wing et al., 2020b), France, the JPL-Table Mountain ~~Facility (TMF) (Ferrare et al., 1995) , California,~~ Observatory Facility (TMO) (Ferrare et al., 1995) and Mauna Loa Observatory (MLO) (Li et al., 2008), ~~Hawaii~~US. All four instruments have very similar power and performance specifications, and follow a similar mode of operation (a few hours per night, 1 to 4 times per week, weather permitting), making it easier to include into a consistent ground-based reference combined dataset. For these instruments, the temperature

total uncertainty ranges from 2-3 K at 30 km to less than 1-2 K between 35 and 55 km, and then back up to 2-5 K in the mid-mesosphere, and up to 20 K near the initialization altitude (80-95 km).

Validating Rayleigh temperature lidar measurements in the upper stratosphere and mesosphere can be difficult due to ~~a~~ lack of reference temperature observations at these altitudes. Occasional comparisons with rocketsonde measurements showed temperature differences of 2-5 K in the lower mesosphere ~~(e. g., Schmidlin (1981); Ferrare et al. (1995)).~~ (Schmidlin, 1981; Ferrare et al., 1995). Over the past two decades, the performance of Rayleigh temperature lidars has been evaluated mainly by comparison with satellite measurements during which they served as the ground-based reference. These inter-comparisons typically yield lidar-satellite differences not exceeding 2-4 K between 30 and 60 km ~~(e. g. Wang et al. (1992); Ferrare et al. (1995); Wu et al. (2003); Sica et al.).~~ (Wang et al., 1992; Ferrare et al., 1995; Wu et al., 2003; Sica et al., 2008; García-Comas et al., 2014; Stiller et al., 2012). At the bottom end of the profiles (below 30-35 km), the lidars have been compared to radiosonde measurements ~~(e. g. Ferrare et al. (1995)).~~ (Ferrare et al., 1995). In the presence of aerosols, Rayleigh backscatter lidars yield a significant cold bias (the thicker the aerosols, the colder the bias). The TMF and MLO instruments include an additional vibrational Raman backscatter channel, i.e., unimpacted by aerosol backscatter, which allows temperature retrieval down to 10 km (Gross et al., 1997; Leblanc et al., 2016), but with a remaining slight cold bias (1-4 K) due to aerosol extinction. For these two lidars the reanalysis packages described in section 2.3 will be compared for greater altitude ranges than the HOH and OHP lidars.

To fill the need for additional validation, one method frequently used within NDACC is to deploy a mobile lidar from site to site to blind-test the lidar instruments permanently deployed at these locations. One ~~of such "traveling standards~~such "travelling standard" is operated by the NASA Goddard Space Flight Centre (GSFC) (McGee et al., 1995) and has been used to validate the lidar instruments at HOH, OHP, TMF, and MLO ~~(Ferrare et al., 1995; Leblanc et al., 1998; Keckhut et al., 2004)~~ (Ferrare et al., 1995; Leblanc et al., 1998; Keckhut et al., 2004; Wing et al., 2020a, b). During these campaigns, temperature differences between the lidars did not exceed 4 K in the 25-60 km altitude range, with typical differences of +/-2 K.

2.2 Microwave Limb Sounder

The NASA EOS MLS was launched on the 14th July 2004 and became operational on the 14th August 2004. It works by observing millimetre and sub-millimetre thermal emission along the limb of the atmosphere. It is in a low polar orbit allowing it to orbit the ~~earth~~ Earth 15 times a day, crossing the equator near to local noon and midnight. It is designed to measure several atmospheric gases and aerosols in the upper troposphere, stratosphere and mesosphere (Waters et al., 2006). It uses the emission from oxygen to provide temperature and pressure measurements; the precision of the measurement is given in Waters et al. (2006) to be 0.5-1 K between 300 and 0.001 hPa, where it has a vertical resolution between 4-8 km. Initial comparisons by Froidevaux et al. (2006) with other satellite retrievals of temperature found that EOS MLS had a 1-2 K warm bias. A more thorough comparison made by Schwartz et al. (2008) compared EOS MLS temperature retrievals to those from radiosondes ~~,~~ and several satellites, ~~and from~~ including GPSRO. It was found that from 0.0001 hPa to 0.3 ~~hPa that~~ hPa the temperature bias could range from -9 to 0K with temperature precision ranging from $\pm$1 to $\pm$2.5 K. A further study by Wing et al. (2018) found that the bias in wintertime MLS was ~~$\pm$10~~-10 K and $\pm$4 K in the summertime. At 1 hPa warm biases from 0 to 5K were found. From 3.16 hPa down to 316 hPa precision was found to be less than $\pm$1 K and biases were between -2 and 1.5 K. Thus,

at pressure heights of 3 hPa and below the EOS MLS satellite has a similar bias to that of the temperature lidar at the same observing height. Wing et al. (2020b) compared EOS MLS at the OHP lidar against NASA's reference lidar and it was found to have a large cold bias above 3 hPa of -10 K. Figure 1 panels a-d shows the mean MLS and Lidar temperature profiles between 2004 and 2017 for HOH, MLO, OHP and TMO respectively. Panels e-h shows the average difference between the matched profiles at HOH, MLO, OHP and TMO respectively. This shows a cold bias which increases in magnitude with height and agrees well with the findings of Schwartz et al. (2008),Wing et al. (2018) and Wing et al. (2020b).

2.3 European Centre for Medium Range ~~Forecasting~~ Forecasts data

A reanalysis dataset is generated by combining many different historical measurements using data assimilation to create an accurate numerical representation of the Earth's atmosphere at a given time. ERA interim spans from 1979 to ~~present and at the time of writing runs 2-3 months behind real time. It is described in Dee et al. (2011) and~~ 2019. ERA-interim has 60 model levels spanning the surface to 0.3 hPa (57 km altitude) with an approximate 79 km horizontal grid resolution and 6 hour analysis windows ~~and~~ (Dee et al., 2011) . It is based on the Integrated Forecasting System (IFS) cycle 31R2 and utilises a 4Dvar data assimilation system. Dee et al. (2011) state that during December 2006 GPSRO data from the COSMIC constellation was included in reanalysis datasets. (Poli et al., 2010) showed that the variability in ERA-interim temperature was much improved after this time. ~~Due to the temporal range of the ERA-interim reanalysis we are able to compare the entire temperature lidar datasetsshown in table 1.~~

~~The ERA-5~~ ERA5 is the 5th generation reanalysis created by the ECMWF and replaces ERA-interim. The new ~~ERA-5~~ ERA5 reanalysis, based on IFS cycle ~~41r2 (ECMWF, 2016)~~ 41R2 (Hersbach et al., 2020), has 137 model levels from ~~the~~ surface to 0.01 hPa (approximately 80 km) and a global horizontal resolution of 31 km, compared to ERA Interim's 60 model levels and 79 km horizontal resolution. ~~Other improved features~~ Amongst 10 years of research and development due to the constant evolution of the ECMWF IFS are the inclusion of more ~~measurements and~~ measurement systems, improved bias correction techniques ~~to assimilate them and better climate forcings(Hersbach and Dee, 2016). Dee et al. (2011) state that during December 2006 GPSRO data from the COSMIC constellation was included in reanalysis datasets and that the variability in these datasets was much improved after this time (Poli et al., 2010). For the second part of the analysis, comparisons with both ERA-interim~~ and ~~ERA-5 before and after the inclusion of COSMIC GPSRO data will be undertaken. In a previous study by Poli et al. (2010) the comparison was made between 200 hPa and 100 hPa. Here the comparison will be made at 100 hPaand above~~model physics, CMIP5 radiative forcings, and data assimilation using a hybrid 4Dvar system (Hersbach et al., 2020) . Simmons et al. (2020) showed that temperature bias in the upper stratosphere of ERA5 was significantly effected with the addition of AMSU-A data between 2000 and 2007 at heights above 15 hPa.

3 ERA-interim comparisons

In this section MLS and lidar profiles will be compared with temperature profiles from ERA-interim. ~~For all of our comparisons we will only focus on the Northern Hemisphere winter months between October and March. Stratospheric variability increases~~

during these months and we examine whether this variability is present within the reanalyses. The lidar temperature profiles were interpolated onto ERA-interim's model levels using geopotential height Z, for time steps that were closest to the mid-point of the lidar's profile acquisition period. Due to differences of To ensure the comparison is accurate the lidar's height coordinate which is in geometric height was first converted to geopotential height. Despite being similar near ground level, the differences are between 0.4 km to and 1 km between geometric height, the height co-ordinate used by the lidars, and Z at the altitude range for the altitude ranges used in this studythe lidar's geometric height was converted to Z prior to the interpolation. The height . The vertical resolution for ERA-interim at these heights is approximately 1.5 km at 10 hPa and 3 km at 1 hPa. For this comparison we use lidar and reanalysis profiles for the full temporal lidar temperature profile range shown in table 1. Figure 2 panels (a-d) shows matched mean temperature profiles from both lidar (red) and ERA-interim (blue) for the lidar sites at Hohenpeissenburg, Mauna-LoaHOH, MLO, OHP and Table Mountain Observatory TMO respectively. Panels e-h show the mean of the matched differences for the corresponding lidars profiles (a-d); grey shading shows 1 standard deviation in the matched differences. ERA-interim at the points studied here has a cold bias in the region of -3 to -4 K. Red dots show Crosses are red where the mean difference is significantly different from zero at the 95% significance level. For HohenpeissenburgHOH, OHP and Table Mountain ObservatoryTMO, the peak cold bias is centred between 10hPa and 1hPa. For Mauna Loa MLO the cold bias extends down to the 100 hPa pressure surface. Reasons for this could be due to Hohenpeissenburg and Table Mountain Observatory being at higher latitudes , whereas at TMO the cold bias is much closer to zero between 10 hPa and 100 hPa. A possible hypothesis is that TMO is at a higher latitude than the tropically positioned Mauna-Loa station where the structure MLO where the representation of the middle atmosphere within ERA-interim differs slightly. In all cases above For all sites ERA-interim exhibits a warm bias between 1 hPa ERA-interim is much warmer as it approaches and 0.1 hPa. This contrastingly warm bias is due to the model top being reached and the stratopause not being represented. Panels i-l show the seasonal variation of the temperature biases with height. The warm bias near the model top is present through out the year, the cold bias between 1 hPa and 10 hPa is present throughout the year, which intensifies at all sites during the November to February period.

A similar analysis was performed using the EOS MLS data. The EOS MLS temperature data was first sorted so that only night time passes within 2.5^o of each temperature lidar site were retained. The remaining EOS MLS temperature profiles were interpolated on to onto ERA-interim's model levels using Z from the reanalysis package for time steps that were closest and less than 3 hours apart. Due to EOS MLS only being launched in 2004 the results of the ERA-interim and MLS comparison shown in figure 3 are for the years 2004 to 2017. Figure 3 shows matched mean temperature profiles from both MLS (red) and ERA-interim (blue) for lidar sites at Hohenpeissenburg, Mauna-LoaHOH, MLO, OHP and Table Mountain Observatory TMO respectively. Panels (e-ge-h) show the mean of the matched differences for MLS over the 4 lidar sites. The cold bias seen in figure 2 between is not present when comparing with EOS MLS, instead a warm bias is present of approximately 1 hPa and 10 hPa is still present. However the size of the bias is smaller with K from 100 hPa to 1 hPa at HOH, MLO and OHP. At TMO a cold bias of -1 to -2 K . For comparisons over Hohenpeissenburg and OHP (panels e and g) the cold bias detected by the MLS is reduced when compared is observed at 1 hPa, the large warm bias above 1 hPa near the ERA-interim model top shown in figure 2 is also observed. MLS shows a different temperature bias to that of the temperature lidar. In figure 1 we see that MLS has a

cold bias when compared to the lidar ~~. Between 10 hPa and 100 hPa MLS observes a warm bias of 1 K in~~ at HOH and OHP and also in Schwartz et al. (2008), Wing et al. (2018) and Wing et al. (2020b). As the MLS records a negative bias when compared with the Lidar, and ERA-interim ~~which cannot be verified by the temperature lidar at these two locations. For comparisons with Table mountain observatory and Mauna-Loa~~ also exhibits a cold bias ~~is also observed from 1 hPa down to 100 hPa similar to the bias found by the lidar at these two sites. The only difference is that the magnitude of the bias is reduced by 1-2 K. This may be~~ compared to the lidar, we see a resulting warm bias in the ERA-interim comparison with MLS. The cold bias seen at 1 hPa at TMO is due to the warm bias ~~of EOS MLS at these heights as shown in Schwartz et al. (2008). This~~ in MLS. The results seen here further demonstrates evidence of a systematic bias between the lidar and MLS measurement technologies ~~.~~ at these heights. There is some oscillatory structure in the temperature bias that has been seen in Wing et al. (2018) who compared EOS MLS to lidar, and found that the oscillations are not retrieval level dependent. Furthermore the amplitude of the oscillations falls within the precision of EOS MLS given in section 2.2

Simmons et al. (2020) compared the global mean temperature from ERA-interim with global mean radiosonde data between 15 hPa and 7 hPa, and 7 hPa and above. Here our results in figure 2 agree with a cold bias seen in Simmons et al. (2020) ~~'s results at 7 hPaand above~~for the height range 15-7 hPa, although the magnitude is much smaller. For the ~~15 hPa to~~ 7 hPa range ~~, Simmons et al. (2020)~~ and above our results disagree with Simmons et al. (2020) who found that there was a ~~slight~~ warm bias of ~~0.5 K . Figure 2 panels (e) and (h) and figure 3 panels (e, g and h) show evidence towards this~~ 1-2 K whereas our results using the lidar a show a cold bias at these heights at HOH , MLO and TMO. The temperature bias shown in Simmons et al. (2020) are a global mean, whereas our measurements are made at fixed locations, Moreover, the bias is calculated using radiosonde data which are not independent as they are assimilated into ERA-interim. This explains the polarity difference in the results shown in this study.

In brief conclusion~~two independent observation methods~~ , the temperature lidars have shown that a cold bias of -3 to -4 K exists between 1 and 10 hPa in ERA-interim ~~over Hohenpeissenburg~~ for the HOH and OHP lidar sites and between 1 and 100 hPa for the ~~Table Mountain observatory and Manua-Loa sites. The reason for the difference in the magnitude of the bias is likely down to the differences in measurement technology. As discussed in section 2, the MLS has been found to exhibit a warm bias between 1 and 3 hPa (Schwartz et al., 2008), which may be the reason for a smaller cold bias than detected with the lidar.~~ TMO and MLO sites.

4 ~~ERA-5 Comparisons~~ERA5 comparisons

To compare the temperature biases in ~~ERA-5~~ ERA5 with those in ERA-Interim we repeat the comparison for ERA5 in a similar manner. The temperature lidar data was interpolated, accounting for geometric height, onto ~~ERA-5~~ ERA5 model levels using Z for the nearest analysis window to the mid point of the lidar acquisition window. The vertical resolution at 10 hPa is 750 m and at 1 hPa is 1.6 km. Figure 4 panels (a-d) show the mean temperature profiles ~~of~~ for the lidar in red and ~~ERA-5~~ ERA5 in blue up to a height of 0.5 hPa for the ~~winter months, October-March for the~~ period 1990-2017. At first inspection ERA5 profiles track the lidar profiles more closely than those of ERA-interim~~(~~, including a more accurate representation of the stratopause

than that of ERA-interim (see figure 2). Figure 4 panels (e-h) show the mean of matched differences with height; ~~red dots show when~~ grey shading shows 1 standard deviation in the matched differences and crosses are red where the mean difference is different from zero at the 95% significance level inferred by a single sample t-test. The temperature biases are significantly smaller than in ERA-interim. For the ~~Mauna-Loa and Table mountain~~ MLO and TMO sites the temperature bias is very close to zero up to the ~~3~~10 hPa pressure surface. ~~Above this height~~ At 10 hPa to 5 hPa a cold bias ~~, which increased to -5~~of -1 to -2 K is observed at all ~~four sites. Hohenpeissenburg shows smaller temperature biases which vary by~~ sites. From 5 hPa to 1 ~~K about zero. OHP shows a smaller varying cool bias about -2 K up to 3 hPa~~ hPa the bias drops to near zero again and above 1 hPa a -3 K cold bias is observed at all four sites. When we consider that the measurement accuracy of the lidar is $\pm$ 2 K, ERA5 gives a good thermal representation of the atmosphere up to ~~3~~1 hPa. Figure 4 panels (i-l) shows the seasonal variation of temperature bias with height. For all four sites there is a slight warm bias of approximately 1 k at 1-5 hPa during the summer months (May-August) with the exception of MLO where the warm bias at this height spans January to July. The cold bias above 1 hPa intensifies in the winter months.

Figure 5 shows the temperature comparisons against MLS for the period 2004 to 2017; the data was interpolated ~~on to~~ onto the ERA5 model ~~data~~ levels using the same method as discussed in section 3. MLS and ~~ERA-5 show good~~ ERA5 show a fair agreement at all sites ~~at all heights. Figure 5 panels (e-h) show that at the Hohenpeissenburg and OHP site~~ between 100 hPa and 10 hPa. But not as good as when compared to that of the lidar at TMO and MLO. From 100 hPa to 5 hPa there is a ~~slight warm bias , which peaks at 1 Kat 3 hPa. The temperature bias at Table Mountain Observatory (g) varies about zero but~~ warm bias which varies with amplitude between 1-2 K. The warm bias peaks at 3 hPa ~~similar to the OHP and Hohenpeissenburg sites. The temperature bias at the Mauna Loa observatory in (f) varies about zero by approximately 0.5 K over the height range studied. The uncertainties in the bias are larger for MLSat higher altitudes which is shown by the increased breadth in the shading in figure 5 panels (e-h). This is likely~~ with an amplitude of 3 K, at 1 hPa and above a large cold bias of -5 K is found. Panels (i-l) show that there is little seasonal variation in the temperature bias between ERA5 and MLS. Given the findings of Schwartz et al. (2008), Wing et al. (2018) and figure 1 it is clear that these bias are largely due to the ~~accuracies of MLS previously discussed. In addition, MLS was found to have a warm bias (Schwartz et al., 2008) between 0.3 hPa and 3 hPa which is seen here. Further to this MLS shows an oscillation with heights that is not seen in the lidar data.~~ MLS temperature bias previously discussed in section 2.2.

Simmons et al. (2020) compared the global mean temperature from ~~ERA-5~~ ERA5 with global mean radiosonde data from 15 hPa upwards and found that performance was similar to ERA-interim pre 2000. A ~~cold bias of -2~~warm bias of 2 K was found above 7 hPa and a slight ~~warm~~ cold bias of less than ~~0.5~~-0.5 K between ~~7-15 hPa for ERA-5~~7 hPa and 15 hPa for ERA5. Between 2000 and 2007 there was an increase in the ~~cold~~ warm bias making the bias ~~-3~~3 K and ~~-0.5~~0.5 K for the layers 7 hPa and above and 7-15 hPa respectively. Simmons et al. (2020) believed that this may be due to the introduction of observations from the AMSU-A ~~satellite (Aumann et al., 2003) data stream at the beginning of the period.~~ instrument aboard NOAA 15 and NOAA 16 (Aumann et al., 2003) from 2000. Post 2007, after the introduction of GPSRO data, Simmons et al. (2020) ~~shows~~ showed the temperature bias above 7 hPa to be ~~-1.5~~1.5 K and less than ~~-0.5~~0.5 K between 7-15 hPa. As our comparisons using the lidar span 1990 to 2017 we do see some aspects of ~~Simmons et al. (2020)'s results~~ findings in Simmons et al. (2020) in

that a ~~cold~~ warm bias is observed above 7 hPa ~~and the temperature bias is small in the 7-15 hPa range. However, due to our methodology of averaging over 30 years for the lidar data it is hard to see if the increase in bias during 2000-2007 is present. Due to the time period of the MLS such a comparison is not possible~~

~~MLS and ERA-5 temperatures agree well over the height ranges studied. The temperature lidar agrees up to 3 hPa with a cold bias observed above this. One reason MLS may not detect this cold bias is due to the fact it exhibits a warm bias, counteracting the cold bias at heights above 3 hPa (Schwartz et al., 2008). Aside from the two differing measurement techniques, the main difference in this study is that the lidar dataset is almost double the length of the EOS MLS dataset- temperature lidars have been used since the 1990s and EOS MLS only became operational in 2004. During this time period other measurement technologies have been added to those already assimilated by reanalysis datasets. It could be that the cold bias observed by the~~ during the summer months. In this section we have shown that ERA5 gives a much improved representation of the upper stratosphere when compared to ERA-interim. The focus has been on calculating temperature bias over a 1990-2017 and 2004-2017 study period for the lidar ~~above 3 hPa in ERA-5 is from an earlier time period in the dataset where observations were more sparse. We check this in the next section by examining temperature bias inferred by the lidar over three different time periods.~~ and MLS respectively. This means it is difficult to see if a particular data stream improved ERA-interim or ERA5's representation of the upper stratosphere. Given that the lidars have made measurements for over 25 years we can examine the data further by year to see how the introduction of new observation streams such as GPSRO and AMSU-A affected the ERA datasets.

~~5 ERA performance due to Assimilation of Satellite data~~

~~In December 2006 the COSMIC constellation was launched and later that month GPSRO data was made available to assimilate into the ECMWF IFS models (Dee et al., 2011; Poli et al., 2010). Initial comparisons by~~

5 ERA performance due to assimilation of COSMIC GPRSO and AMSU-A

The mean of differences between lidar and both ERA-interim and ERA5 were decomposed by year to examine if the introduction of new satellite data streams such as COSMIC GPSRO and AMSU-A changed the stratospheric temperature bias. Poli et al. (2010) showed that the inclusion of COSMIC GPSRO data improved ~~model performance~~ reanalysis bias in the upper troposphere and lower stratosphere. However~~this~~ , their comparison was only undertaken between 200 hPa and 100 hPa. ~~Its effect~~ The effect of the inclusion of COSMIC GPSRO at 100 hPa was described in Cardinali and Healy (2014) who found that the introduction of COSMIC GPSRO showed ~~signs of improvement at higher levels. Also~~ a decrease in forecast error between 250 hPa and 50 hPa. Simmons et al. (2020) showed a change in ~~ERA-5 temperature bias at~~ ERA5 temperature bias after the inclusion of the AMSU-A satellite data ~~. Here, an analysis is undertaken which will compare ERA-interim and ERA5 data in 1999/2000. Data from MLS is only available post 2004 and given AMSU-A data became available 19998-2000 we restrict~~ our time series analysis to the lidar data ~~both pre 1st January 2000, 1st January 2000 to 1st January 2007 and post 1st January~~

2007 to see how the introduction of AMSU-A and COSMIC GPSRO data effects upper stratosphere temperatures in ERA-5 and ERA-interimonly.

Figure 6 shows temperature difference the average annual temperature difference as a function of year and height between ERA-interim and the temperature lidarin this study for pre 2000 (green), 2000 to 2007 (purple) and post 2007 (cyan) data. The cold temperature bias at Hohenheissenburg (a) and OHP (c) in ERA-interim does not seem to affected significantly by either of AMSU-A or COSMIC GPSRO data . At Mauna-Loa between 5-10 hPa there is an increase in the cold . HOH, OHP and TMO show an increase of the cold bias from -1 to -2 K post 2000 which maybe due to the introduction between 1 and 10 hPa until 1995-1996. It is not known if this occurred at MLO, due to lack of data over this period. Figure 14 in Dee et al. (2011) shows that during this 1995/1996 period the NOAA 14 SSU unit was launched and added to the data assimilation streams. This could explain the decrease in the cold bias at the 1 hPa to 10 hPa pressure range. During the 1998/1999 period a warm bias, similar to that experienced at the model top formed around the 1 hPa pressure level, which coincided with the inclusion of AMSU-A satellite . TMO in panel (d) of figure 6 is the only site which shows significant improvement due to the inclusion of AMSU-A and COSMIC GPSRO with a pre 2000 cold bias of -6 K and a post 2007 bias of -3 Kdata in 1998. Simmons et al. (2020) discussed how the addition of AMSU-A may have an affect on ERA-interim. It is apparent here it may also increase the warm-bias of ERA-interim at 1 hPa. HOH and TMO show subtle reductions in the cold bias from -5 to -4 in the 10-1 hPa range post 2007. With the exception of OHP which has an intensifying cold bias between 2014 and 2017 there are no further significant changes in temperature bias over the studied period.

Figure 7 shows the temperature difference between ERA-5 average annual temperature difference as a function of year and height between ERA5 and the temperature lidardata in this study for pre 2000 (green), 2000 to 2007 (purple) and post 2007 (cyan) data. For Hohenpeissenburg, Mauna-loa and OHP the introduction of AMSU-A data appears to have little effect on the temperature bias at these heights. The only significant improvement is a reduction in . In the 3-5 hPa range at all sites there is a warm bias of 2-3 K between 1994 and 1997/1998. Given the abrupt and consistent reduction of the warm bias at all sites during 1998, the year AMSUA aboard NOAA-15 data began being assimilated, it is clear it reduced the warm bias at 5 hPa at Mauna-Loa. Table mountain in panel (d) shows significant differences- the cold bias above 3 hPa is reduced to under -5 K and the cold bias between 5 hPa and 10 hPa has switched to a warm biasof this height. However by 2000 the warm bias returned at this height range and is most dominant at OHP and TMO. One explanation could be that the addition of AMSUA aboard NOAA-16 which was ingested between the years of 2001-2009 (Hersbach et al., 2020) could be the reason for this warm bias, agreeing with the conclusions in Simmons et al. (2020). The cold bias at 1 K, an overall improvement. The temperature bias at TMO is further improved post 2007 with temperature biases of $\pm$ 1.5 K between 1 and 100 hPa . For the Hohenpeissenburg site, although the cold bias has been reduced above 3 hPa for the post 2007 periods, the cold bias in the 3-10 hPa range has increased. There are no significant changes between the 2000 and 2007 and post 2007 values for Mauna-Loa and for OHP the cold bias increases to -2 K post 2007. hPa at HOH, MLO and TMO reduced from -2K to near 0 K at the end of 2006 which coincided with COSMIC GPSRO being made available for assimilation from December 2006. Although GPSRO has a tangent height of 50 km the assimilation of the bending angle means it can contribute observations at higher altitudes than

50 km (Healy, 2008). Additionally the hydrostatic nature of the model means that observations assimilated at a given level affect those above and below.

In summary, the cold temperature bias in ERA5 above ~~3~~1 hPa is reduced from ~~5~~-2 K to ~~1-3~~near 0 K at 3 out of 4 of the sites post 2007 due to the inclusion of GPSRO ~~and~~ data. Inclusion of the AMSU-A data ~~. This demonstrates how a lengthy time set such as the temperature lidar can be used to assess the effect of the addition of different observation streams in reanalysis data sets in the upper atmosphere.~~ on NOAA-15 from 1998 appears to reduce a warm bias at 3 hPa. This warm bias reappears at some sites with the introduction of AMSU-A data from NOAA 16. Although the instruments on both satellites are similar, inter satellite brightness temperature bias has been shown before by Mo (2011) between the AMSU-As on both NOAA-18 and NOAA-19 satellites. This could explain the opposing bias seen here. When comparing these findings to that of Simmons et al. (2020) who showed an intensifying of ~~the cold~~ a warm bias between 2000 and 2007 ~~we see that in these comparisons this is not the case. The reasons for this may potentially be that the metric used in Simmons et al. (2020) was the global mean radiosonde temperature with a bias correction applied. However, the radiosonde data had been assimilated into ERA-5 already and this combined with the bias correction may lead to the cold bias shown. Alternatively and more likely is that as a global average was used, this may not represent temperature bias at individual sites, as is shown here. It is well known that temperature bias across the globe at the heights considered here is not homogeneous.~~ globally we see agreement at OHP and TMO.

6 Conclusions

In this work we have utilised ~~both~~ temperature lidar from the NDACC network~~and EOS MLS data~~, that are not assimilated into ~~models~~reanalyses, to identify temperature bias in the upper ~~atmosphere of ECMWF reanalysis datasets~~stratosphere in ERA-interim and ERA5 reanalyses. For comparisons with ERA-interim and the lidar, a cold bias of -3 to -4 K between 10 hPa and 1 hPa and a large warm bias above 1 hPa was found. ~~Comparisons with MLS also yielded cold biases between 10 hPa and 1 hPa but size of the cold bias was smaller at 2-3 K. Given the warm bias from MLS and that both measurement technologies have an approximate accuracy of ± 2 K a fair assessment of the cold~~ The cold bias ~~in ERA-interim is between -4~~ intensified in Northern hemisphere winter to -5 to ~~-2~~ -6 K. ~~Temperature bias~~ For ERA5 the temperature bias between the lidar and ~~ERA-5 were within~~ ERA5 is within ±1 K to a height of 1 hPa, which given the measurement accuracies of the lidar ± 2 K gives a good thermal representation of the stratosphere. Above 1 hPa a cold bias of -2 to ~~a height of 3 hPa and MLS showed similar agreement over the whole study height. Comparisons between~~ -3 K is found. Similar to ERA-interim, ERA5 ~~and MLS do hint at a slight warm bias at 3 hPa, although this is likely due to EOS MLS's bias at these height ranges. By splitting the lidar dataset into to three time periods, 1990-2000, 2000-2007 and 2007 to 2017, it was found that the~~ cold bias ~~detected by the lidar~~ above ~~3 hPa was reduced from~~ 1 hPa intensifies to -4 K in the Northern hemisphere winter months and becomes a warm bias in the summer months. When comparing to MLS, ERA-interim exhibited a warm bias of 1 K. ERA5 had a warm bias of 1-2 K up to 5 hPa. Above this height MLS's warm bias at 3 hPa and the large cold bias at 1 hPa, both shown here in figure 1, Schwartz et al. (2008) and Wing et al. (2018) saturate the temperature bias results.

When examining the ERA-interim lidar comparison over 1990-2017 we see a warm bias increase at 1 hPa around 1997/1998, which could be due to the introduction of AMSU-A. There is also a small reduction from -5 K to ~~2-3 K with the introduction of GPSRO in 2007.~~ -4 K post 2006 which is most noticeable at 1-2 hPa at HOH and TMO. This could be an indication that the inclusion of COSMIC GPRSO has an affect on the upper stratosphere within ERA-interim. For ERA5 the effects of new satellites being assimilated is clearer. We see that the inclusion of the two AMSU-A data streams had an effect on temperature bias between 1-10 hPa. It appears that in 1998 during the assimilation of AMSU-A from NOAA-15 the warm bias improved. Yet when AMSU-A from NOAA-16 was assimilated in 2000-2009 the warm bias returned. However it was not as intense. Post 2007 a reduction in the cold bias above 1 hPa was observed at HOH, MLO and TMO due to the assimilation of COSMIC GPSRO. Other small changes in the temperature bias seen in figures 6 and 7 could be attributed to other observations being assimilated. Both Hersbach et al. (2020) and Dee et al. (2011) show that in both ERA5 and ERA-interim the amounts and type of observations increase with time making it harder to characterise behaviours in the upper stratospheric temperature bias to a particular observation dataset.

From the comparisons here it can be stated that ~~ERA-5~~ ERA5 is much improved compared to ERA-interim and has a good thermodynamic representation of the upper stratosphere~~, when~~ . When we consider the uncertainties in the lidar ~~and MLS at these heights are ± 2 K, making it~~ ERA5 is an excellent choice for further stratospheric studies or ~~using it as reference with which~~ for the use as reference to compare other ~~models~~reanalyses to. However, a cold bias detected in ~~ERA-5~~ ERA5 by the lidar before the inclusion of GPSRO and AMSU-A data should be accounted for ~~.~~ in studies such as Shangguan et al. (2019) and Bohlinger et al. (2014). These studies use both ERA5 and ERA-interim to assess long term and short term stratospheric temperature variability. In future works exploring stratospheric temperature trends, changes in temperature biases presented here need to be considered.

~~Using both temperature lidar and a non-assimilated satellite such as EOS MLS, both have their advantages. EOS MLS has global coverage albeit coarser vertical resolution, and observations over a shorter time period.~~ The temperature lidars~~on the other hand~~, whilst limited to a few locations globally, have high vertical resolution measurements that have been made for nearly 30 years, making them an important and useful tool for inferring temperature ~~bias~~ biases in reanalysis datasets, which span similar time periods. A future test could see the lidar networks used to explore modifications to reanalysis datasets such as testing the experimental ERA5.1 ~~shown in Simmons et al. (2020)~~discussed in Simmons et al. (2020).

Data availability. The temperature lidar data is available for public download through http://ndacc-lidar.org/index.php?id=70/Data.htm. The Microwave Limb Sounder data was available for public download at https://mls.jpl.nasa.gov/. ECMWF ERA-interim and ERA5 data are available from the ECMWF MARS archive.

Author contributions. GJM extracted the datasets, performed the analysis and prepared the manuscript, RW provided and processed the OHP data, ACP, RGH, IP, AH and PK provided inputs into the analysis and preparation of the manuscript, TB provided MLO and TMO data and assisted with manuscript preparation, WS provided the HOH data and assisted with manuscript preparation.

Competing interests. There are no competing interests

Acknowledgements. This work was performed during the course of the ARISE2 collaborative infrastructure design study project funded by the European Commission H2020 program (grant number 653980, www.arise-project.eu).The authors also wish to acknowledge staff at the ECMWF for their discussions as this work advanced.

References

Aumann, H. H., Chahine, M. T., Gautier, C., Goldberg, M. D., Kalnay, E., McMillin, L. M., Revercomb, H., Rosenkranz, P. W., Smith, W. L., Staelin, D. H., et al.: AIRS/AMSU/HSB on the Aqua mission: Design, science objectives, data products, and processing systems, IEEE Transactions on Geoscience and Remote Sensing, 41, 253–264, 2003.

Baldwin, M., Gray, L., Dunkerton, T., Hamilton, K., Haynes, P., Randel, W., Holton, J., Alexander, M., Hirota, I., Horinouchi, T., et al.: The quasi-biennial oscillation, Reviews of Geophysics, 39, 179–229, 2001.

Bohlinger, P., Sinnhuber, B.-M., Ruhnke, R., and Kirner, O.: Radiative and dynamical contributions to past and future Arctic stratospheric temperature trends, Atmospheric Chemistry and Physics, 14, 1679–1688, 2014.

Butler, A. H., Seidel, D. J., Hardiman, S. C., Butchart, N., Birner, T., and Match, A.: Defining sudden stratospheric warmings, Bulletin of the American Meteorological Society, 96, 1913–1928, 2015.

Cardinali, C. and Healy, S.: Impact of GPS radio occultation measurements in the ECMWF system using adjoint-based diagnostics, Quarterly Journal of the Royal Meteorological Society, 140, 2315–2320, 2014.

Chandra, S., Fleming, E. L., Schoeberl, M. R., and Barnett, J. J.: Monthly mean global climatology of temperature, wind, geopotential height and pressure for 0–120 km, Advances in Space Research, 10, 3–12, https://doi.org/https://doi.org/10.1016/0273-1177(90)90230-W, https://www.sciencedirect.com/science/article/pii/027311779090230W, 1990.

Charlton, A. J. and Polvani, L. M.: A new look at stratospheric sudden warmings. Part I: Climatology and modeling benchmarks, Journal of Climate, 20, 449–469, 2007.

Dee, D., Uppala, S., Simmons, A., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M., Balsamo, G., Bauer, P., et al.: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system, Quarterly Journal of the Royal Meteorological Society, 137, 553–597, 2011.

Domeisen, D. I.: Estimating the frequency of sudden stratospheric warming events from surface observations of the North Atlantic Oscillation, Journal of Geophysical Research: Atmospheres, 124, 3180–3194, 2019.

Domeisen, D. I., Garfinkel, C. I., and Butler, A. H.: The teleconnection of El Niño Southern Oscillation to the stratosphere, Reviews of Geophysics, 57, 5–47, 2019a.

ECMWF: Part III: Dynamics and Numerical Procedures, https://www.ecmwf.int/en/elibrary/16647-ifs-documentation-cy41r2-part-iii-dynamics-and-num 2016.

Ferrare, R., McGee, T., Whiteman, D., Burris, J., Owens, M., Butler, J., Barnes, R., Schmidlin, F., Komhyr, W., Wang, P., et al.: Lidar measurements of stratospheric temperature during STOIC, Journal of Geophysical Research: Atmospheres, 100, 9303–9312, 1995.

Froidevaux, L., Livesey, N. J., Read, W. G., Jiang, Y. B., Jimenez, C., Filipiak, M. J., Schwartz, M. J., Santee, M. L., Pumphrey, H. C., Jiang, J. H., et al.: Early validation analyses of atmospheric profiles from EOS MLS on the Aura satellite, IEEE Transactions on Geoscience and Remote Sensing, 44, 1106–1121, 2006.

Fujiwara, M., Wright, J. S., Manney, G. L., Gray, L. J., Anstey, J., Birner, T., Davis, S., Gerber, E. P., Harvey, V. L., Hegglin, M. I., et al.: Introduction to the SPARC Reanalysis Intercomparison Project (S-RIP) and overview of the reanalysis systems, Atmospheric Chemistry and Physics, 17, 1417–1452, 2017.

García-Comas, M., Funke, B., Gardini, A., López-Puertas, M., Jurado-Navarro, A., von Clarmann, T., Stiller, G., Kiefer, M., Boone, C., Leblanc, T., et al.: MIPAS temperature from the stratosphere to the lower thermosphere: comparison of version vM21 with ACE-FTS, MLS, OSIRIS, SABER, SOFIE and lidar measurements., Atmospheric Measurement Techniques Discussions, 7, 2014.

Gross, M. R., McGee, T. J., Ferrare, R. A., Singh, U. N., and Kimvilakani, P.: Temperature measurements made with a combined Rayleigh–Mie and Raman lidar, Applied Optics, 36, 5987–5995, 1997.

Hauchecorne, A.: Lidar Temperature Measurements in the Middle Atmosphere, The Review of Laser Engineering, 23, 119–123, 1995.

Hauchecorne, A. and Chanin, M.-L.: Density and temperature profiles obtained by lidar between 35 and 70 km, Geophysical Research Letters, 7, 565–568, 1980.

Healy, S.: Assimilation of GPS radio occultation measurements at ECMWF, in: Proceedings of the GRAS SAF Workshop on Applications of GPSRO measurements, ECMWF, Reading, UK, pp. 16–18, 2008.

Hersbach, H. and Dee, D.: ERA5 reanalysis is in production, ECMWF Newsletter, 147, 2016.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., et al.: The ERA5 global reanalysis, Quarterly Journal of the Royal Meteorological Society, 146, 1999–2049, 2020.

Jenkins, D., Wareing, D., Thomas, L., and Vaughan, G.: Upper stratospheric and mesospheric temperatures derived from lidar observations at Aberystwyth, Journal of Atmospheric and Terrestrial Physics, 49, 287–298, 1987.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., et al.: The NCEP/NCAR 40-year reanalysis project, Bulletin of the American Meteorological Society, 77, 437–471, 1996.

Keckhut, P., McDermid, S., Swart, D., McGee, T., Godin-Beekmann, S., Adriani, A., Barnes, J., Baray, J.-L., Bencherif, H., Claude, H., et al.: Review of ozone and temperature lidar validations performed within the framework of the Network for the Detection of Stratospheric Change, Journal of Environmental Monitoring, 6, 721–733, 2004.

Kuo, Y.-H., Sokolovskiy, S. V., Anthes, R. A., and Vandenberghe, F.: Assimilation of GPS radio occultation data for numerical weather prediction, Terrestrial Atmospheric and Oceanic Sciences, 11, 157–186, 2000.

Le Pichon, A., Assink, J., Heinrich, P., Blanc, E., Charlton-Perez, A., Lee, C. F., Keckhut, P., Hauchecorne, A., Rüfenacht, R., Kämpfer, N., et al.: Comparison of co-located independent ground-based middle atmospheric wind and temperature measurements with numerical weather prediction models, Journal of Geophysical Research: Atmospheres, 120, 8318–8331, 2015.

Leblanc, T., McDermid, I. S., Hauchecorne, A., and Keckhut, P.: Evaluation of optimization of lidar temperature analysis algorithms using simulated data, Journal of Geophysical Research: Atmospheres, 103, 6177–6187, 1998.

Leblanc, T., Sica, R. J., van Gijsel, J. A. E., Haefele, A., Payen, G., and Liberti, G.: Proposed standardized definitions for vertical resolution and uncertainty in the NDACC lidar ozone and temperature algorithms – Part 3: Temperature uncertainty budget, Atmospheric Measurement Techniques, 9, 4079–4101, https://doi.org/10.5194/amt-9-4079-2016, https://amt.copernicus.org/articles/9/4079/2016/, 2016.

Lee, C., Smets, P., Charlton-Perez, A., Evers, L., Harrison, G., and Marlton, G.: The potential impact of upper stratospheric measurements on sub-seasonal forecasts in the extra-tropics, in: Infrasound Monitoring for Atmospheric Studies, pp. 889–907, Springer, 2019.

Li, T., Leblanc, T., and McDermid, I. S.: Interannual variations of middle atmospheric temperature as measured by the JPL lidar at Mauna Loa Observatory, Hawaii (19.5 N, 155.6 W), Journal of Geophysical Research: Atmospheres, 113, 2008.

McGee, T. J., Ferrare, R. A., Whiteman, D. N., Butler, J. J., Burris, J. F., and Owens, M. A.: Lidar measurements of stratospheric ozone during the STOIC campaign, Journal of Geophysical Research: Atmospheres, 100, 9255–9262, 1995.

Miller, D., Brownscombe, J., Carruthers, G., Pick, D., and Stewart, K.: Operational temperature sounding of the stratosphere, Phil. Trans. R. Soc. Lond. A, 296, 65–71, 1980.

Mo, T.: Calibration of the NOAA AMSU-A radiometers with natural test sites, IEEE transactions on geoscience and remote sensing, 49, 3334–3342, 2011.

Picone, J. M., Hedin, A. E., Drob, D. P., and Aikin, A. C.: NRLMSISE-00 empirical model of the atmosphere: Statistical comparisons and scientific issues, Journal of Geophysical Research: Space Physics, 107, SIA 15–1–SIA 15–16, https://doi.org/10.1029/2002JA009430, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2002JA009430, 2002.

Poli, P., Healy, S., and Dee, D.: Assimilation of Global Positioning System radio occultation data in the ECMWF ERA–Interim reanalysis, Quarterly Journal of the Royal Meteorological Society, 136, 1972–1990, 2010.

Russell, J. M., Mlynczak, M. G., Gordley, L. L., Tansock, J. J., and Esplin, R. W.: Overview of the SABER experiment and preliminary calibration results, in: Optical Spectroscopic Techniques and Instrumentation for Atmospheric and Space Research III, vol. 3756, pp. 277–289, International Society for Optics and Photonics, 1999.

Schmidlin, F. J.: Repeatability and measurement uncertainty of the United States meteorological rocketsonde, Journal of Geophysical Research: Oceans, 86, 9599–9603, 1981.

Schwartz, M., Lambert, A., Manney, G., Read, W., Livesey, N., Froidevaux, L., Ao, C., Bernath, P., Boone, C., Cofield, R., et al.: Validation of the Aura Microwave Limb Sounder temperature and geopotential height measurements, Journal of Geophysical Research: Atmospheres, 113, 2008.

Seviour, W. J., Butchart, N., and Hardiman, S. C.: The Brewer–Dobson circulation inferred from ERA-Interim, Quarterly Journal of the Royal Meteorological Society, 138, 878–888, 2012.

Shangguan, M., Wang, W., and Jin, S.: Variability of temperature and ozone in the upper troposphere and lower stratosphere from multi-satellite observations and reanalysis data, Atmospheric Chemistry and Physics, 19, 6659–6679, 2019.

Sheng, Z., Jiang, Y., Wan, L., and Fan, Z.: A study of atmospheric temperature and wind profiles obtained from rocketsondes in the Chinese midlatitude region, Journal of Atmospheric and Oceanic Technology, 32, 722–735, 2015.

Sica, R., Izawa, M., Walker, K., Boone, C., Petelina, S., Argall, P., Bernath, P., Burns, G., Catoire, V., Collins, R., et al.: Validation of the Atmospheric Chemistry Experiment (ACE) version 2.2 temperature using ground-based and space-borne measurements, Atmospheric Chemistry and Physics, 8, 35–62, 2008.

Simmons, A., Soci, C., Nicolas, J., Bell, B., Berrisford, P., Dragani, R., Flemming, J., Haimberger, L., Healy, S., Hersbach, H., Horányi, A., Inness, A., Munoz-Sabater, J., Radu, R., and Schepers, D.: Global stratospheric temperature bias and other stratospheric aspects of ERA5 and ERA5.1, 859, https://doi.org/10.21957/rcxqfmg0, https://www.ecmwf.int/node/19362, 2020.

Skerlak, B., Sprenger, M., and Wernli, H.: A global climatology of stratosphere-troposphere exchange using the ERA-Interim data set from 1979 to 2011, Atmospheric Chemistry and Physics, 14, 913, 2014.

Steinbrecht, W., McGee, T., Twigg, L., Claude, H., Schönenborn, F., Sumnicht, G., and Silbert, D.: Intercomparison of stratospheric ozone and temperature profiles during the October 2005 Hohenpeißenberg Ozone Profiling Experiment (HOPE), Atmospheric Measurement Techniques, 2, 125–145, 2009.

Stiller, G., Kiefer, M., Eckert, E., Von Clarmann, T., Kellmann, S., García-Comas, M., Funke, B., Leblanc, T., Fetzer, E., Froidevaux, L., et al.: Validation of MIPAS IMK/IAA temperature, water vapor, and ozone profiles with MOHAVE-2009 campaign measurements, Atmospheric Measurement Techniques (AMT), 5, 289–320, 2012.

Susskind, J., Barnet, C., Blaisdell, J., Iredell, L., Keita, F., Kouvaris, L., Molnar, G., and Chahine, M.: Accuracy of geophysical parameters derived from Atmospheric Infrared Sounder/Advanced Microwave Sounding Unit as a function of fractional cloud cover, Journal of Geophysical Research: Atmospheres, 111, 2006.

Wang, P.-H., McCormick, M., Chu, W., Lenoble, J., Nagatani, R., Chanin, M., Barnes, R., Schmidlin, F., and Rowland, M.: SAGE II stratospheric density and temperature retrieval experiment, Journal of Geophysical Research: Atmospheres, 97, 843–863, 1992.

Waters, J. W., Froidevaux, L., Harwood, R. S., Jarnot, R. F., Pickett, H. M., Read, W. G., Siegel, P. H., Cofield, R. E., Filipiak, M. J., Flower, D. A., et al.: The earth observing system microwave limb sounder (EOS MLS) on the Aura satellite, IEEE Transactions on Geoscience and Remote Sensing, 44, 1075–1092, 2006.

Wing, R., Hauchecorne, A., Keckhut, P., Godin-Beeckmann, S., Khaykin, S., and Mccullough, E.: Lidar temperature series in the middle atmosphere as a reference data set–Part 2: Assessment of temperature observations from MLS/Aura and SABER/TIMED satellites, Atmospheric Measurement Techniques, 11, 6703–6717, 2018.

Wing, R., Godin-Beekmann, S., Steinbrecht, W., McGee, T. J., Sullivan, J. T., Khaykin, S., Sumnicht, G., and Twigg, L.: Evaluation of the New NDACC Ozone and Temperature Lidar at Hohenpeißenberg and Comparison of Results with Previous NDACC Campaigns, Atmospheric Measurement Techniques Discussions, pp. 1–46, 2020a.

Wing, R., Steinbrecht, W., Godin-Beekmann, S., McGee, T. J., Sullivan, J. T., Sumnicht, G., Ancellet, G., Hauchecorne, A., Khaykin, S., and Keckhut, P.: Intercomparison and evaluation of ground-and satellite-based stratospheric ozone and temperature profiles above Observatoire de Haute-Provence during the Lidar Validation NDACC Experiment (LAVANDE), Atmospheric Measurement Techniques, 13, 5621–5642, 2020b.

Wu, D., Read, W., Shippony, Z., Leblanc, T., Duck, T., Ortland, D., Sica, R., Argall, P., Oberheide, J., Hauchecorne, A., et al.: Mesospheric temperature from UARS MLS: retrieval and validation, Journal of Atmospheric and Solar-Terrestrial Physics, 65, 245–267, 2003.

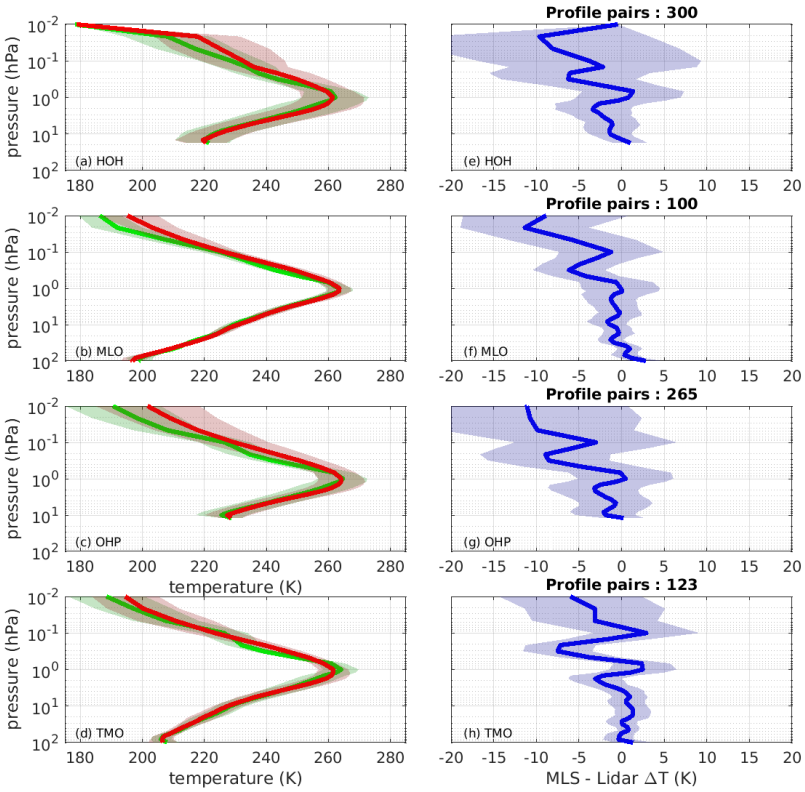

Figure 1. Mean profiles of both temperature from EOS MLS (green) and Rayleigh temperature lidar (red) positioned at a) Hohenpeissenburg, b) Mauna Loa, c) Observatoire de Haute-Provence and d) Table Mountain Observatory. Shading depicts 1 standard deviation in the mean temperature. The vertical profiles of the mean of the differences between the lidar and MLS for the lidar shown in a-b is shown in c-d respectively, shading shows 1 standard deviation within the mean difference.

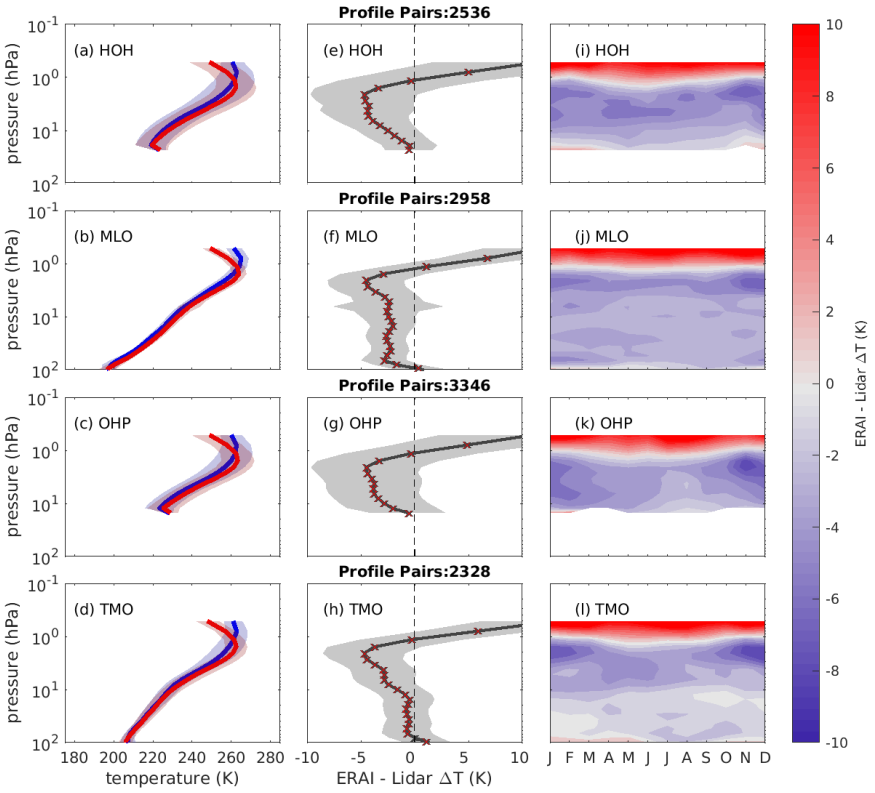

Figure 2. Mean profiles of matched temperatures from ERA-interim (blue) and Rayleigh temperature lidar (red) positioned at a) Hohenpeissenburg, b) Mauna Loa, c) Observatoire de Haute-Provence and d) Table Mountain Observatory ~~for winter months (ONDJFM)~~ between 1990 and 2017. Shading depicts 1 standard deviation in the mean temperature. The vertical profiles of the mean of the differences between ERA-interim and each lidar shown in a-d ~~is~~ are shown in e-h respectively; shading shows 1 standard deviation of mean difference. ~~Red dots show model levels~~ Crosses are red where the mean of the differences were different from zero at the 95% significance level. Panels i-l show the temperature differences as a function of month and pressure level for each of the lidars shown in panels a-d

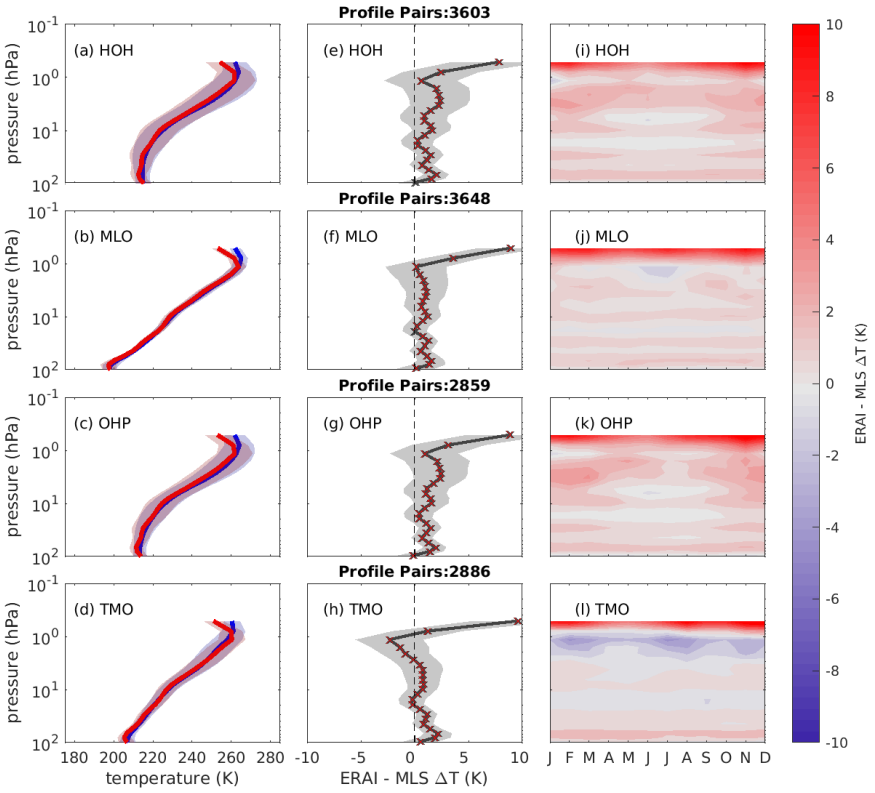

Figure 3. Mean profiles of matched temperatures from ERA-interim (blue) and EOS MLS passes (red) at a) Hohenpeissenburg, b) Mauna Loa, c) Observatoire de Haute-Provence and d) Table Mountain Observatory ~~for winter months (ONDJFM)~~ between ~~2000~~ 1990 and 2017. Shading depicts 1 standard deviation in the mean temperature. The vertical profiles of the mean of the differences between ERA-interim and EOS MLS shown in a-d ~~is~~ are shown in e-h respectively; shading shows 1 standard deviation of mean difference. ~~Red dots show~~ Crosses are red for model levels where the mean of the differences ~~was~~ are different from zero at the 95% significance level.Panels i-k show the temperature differences as a function of month and pressure level for each of the lidars shown in panels a-d.

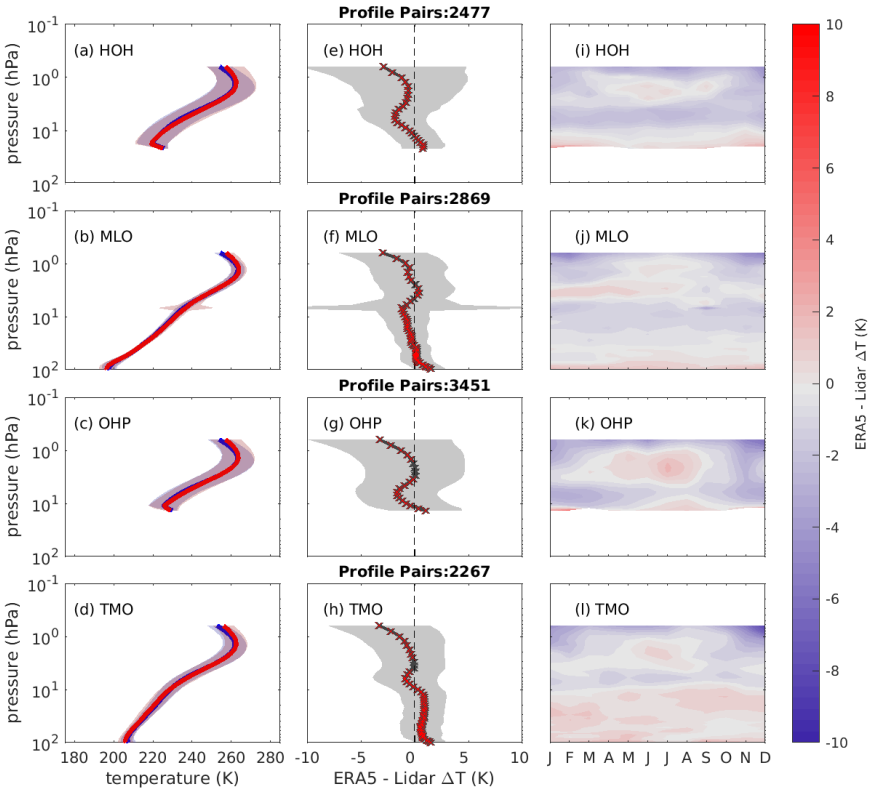

Figure 4. Mean profiles of matched temperatures from ~~ERA-5~~ ERA5 (blue) and Rayleigh temperature lidar (red) positioned at a) Hohenpeissenburg, b) Mauna Loa, c) Observatoire de Haute-Provence and d) Table Mountain Observatory ~~for winter months (ONDJFM)~~ between 1990 and 2017. Shading depicts 1 standard deviation in the mean temperature. The vertical profiles of the mean of the differences between ~~ERA-5~~ ERA5 and each lidar shown in a-d ~~is~~ are shown in e-h respectively; shading shows 1 standard deviation of mean difference. ~~Red dots show~~ Crosses are red for model levels where the mean of the differences ~~were~~ are different from zero at the 95% significance level. Panels i-l show the temperature differences as a function of month and pressure level for each of the lidars shown in panels a-d.

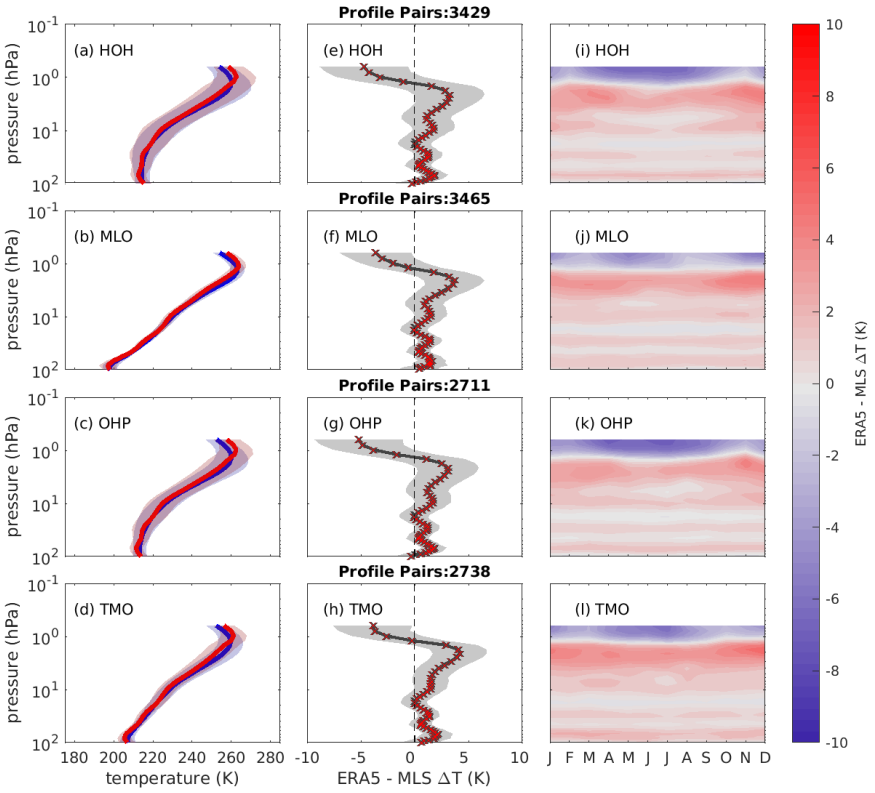

Figure 5. Mean profiles of matched temperatures from ~~ERA-interim~~ ERA5 (blue) and EOS MLS passes (red)at a) Hohenpeissenburg, b) Mauna Loa, c) Observatoire de Haute-Provence and d) Table Mountain Observatory ~~for winter months (ONDJFM)~~ between 1990 and 2017. Shading depicts 1 standard deviation in the mean temperature. The vertical profiles of the mean of the differences between ~~ERA-5~~ ERA5 and EOS MLS are shown in a-d ~~is~~ are shown in e-h respectively; shading shows 1 standard deviation of mean difference. ~~Red dots show~~ Crosses are red for model levels where the mean of the differences ~~was~~ are different from zero at the 95% significance level. Panels i-l show the temperature differences as a function of month and pressure level for each of the lidars shown in panels a-d.

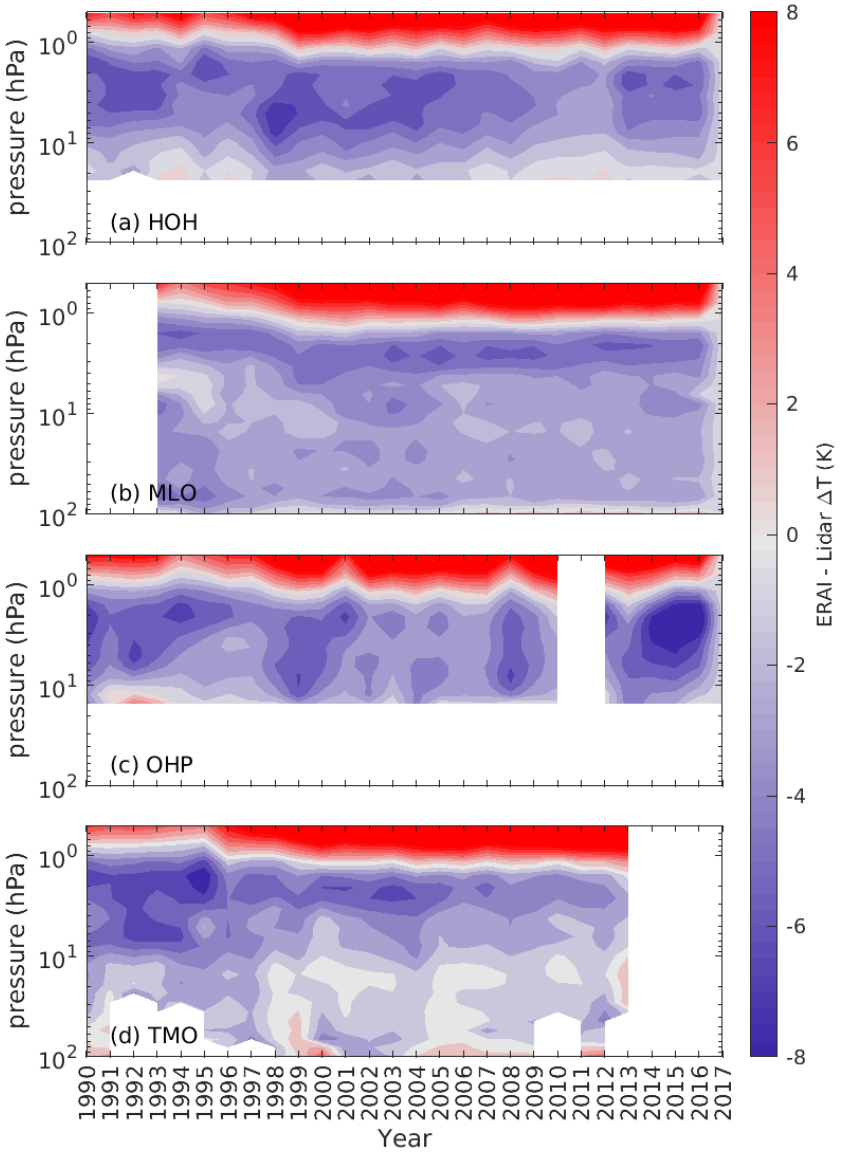

Figure 6. Annual Temperature bias between ERA-interim and temperature lidars at (a) Hohenpeissenburg, (b) Mauna Loa, (c) Observatoire de Haute-Provence and (d) the ~~table mountain observatory for the winter months ONDJFM. Green is ERA-interim comparisons from~~ Table Mountain Observatory plotted as function of year and height between 1990 ~~to 2000, purple is ERA-interim comparisons from 2000 to 2007~~ and ~~blue is ERA-interim comparisons from 2007-2017. The solid lines~~ 2017. Data gaps are ~~the mean and the shading is one standard deviation of the means~~given by white blocks.

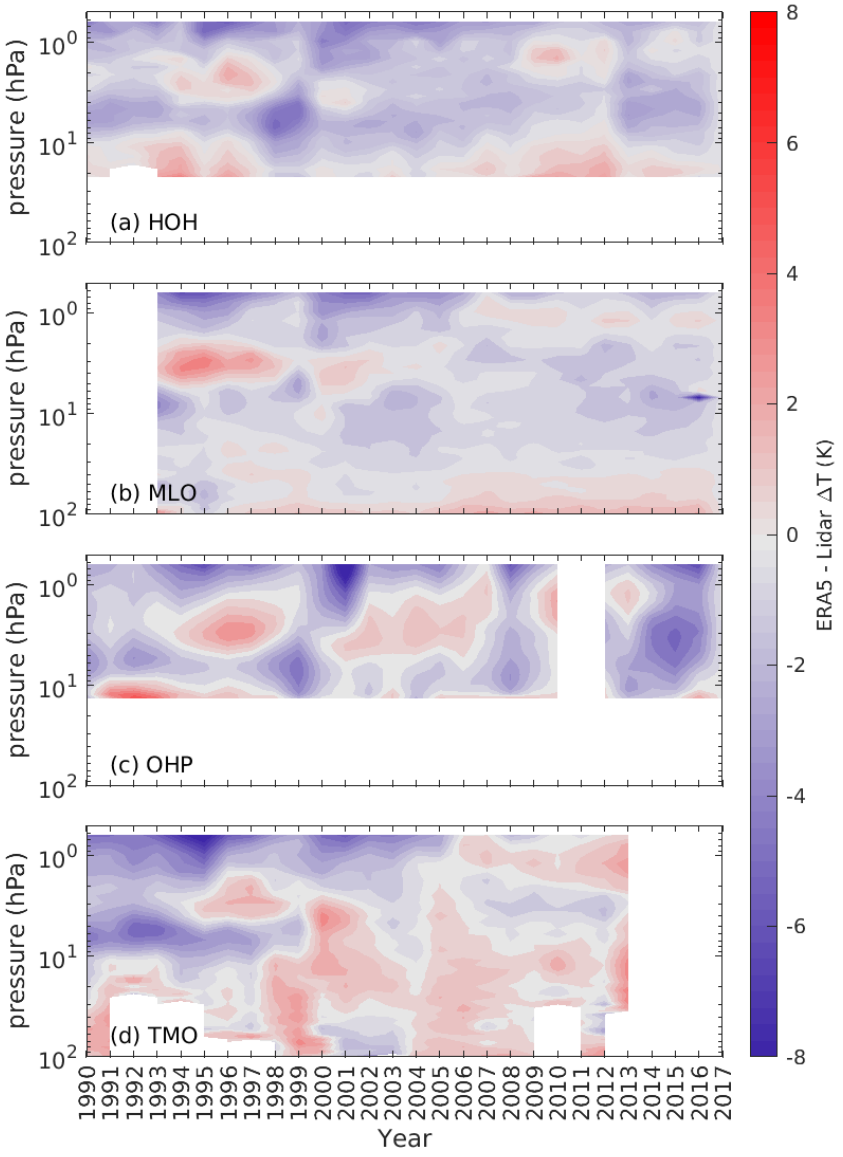

Figure 7. ~~Temperature~~ Annual temperature bias between ~~ERA-5~~ ERA5 and temperature lidars at (a) Hohenpeissenburg, (b) Mauna Loa, (c) Observatoire de Haute-Provence and (d) the ~~table mountain observatory for the winter months ONDJFM. Green is ERA-5 comparisons from~~ Table Mountain Observatory plotted as function of year and height between 1990 ~~to 2000, purple is ERA-5 comparisons from 2000 to 2007~~ and ~~blue is ERA-5 comparisons from 2007-2017. The solid lines~~ 2017. Data gaps are ~~the mean and the shading is one standard deviation of the means.~~given by white blocks

Table 1. Table summarising the geo-spatial and technical information of the 6 NDACC lidars used in this study

Lidar	Lat.^{o}	Lon.^{o}	Period studied[1]	Wavelength (nm)	Range gate Δz (m)
Hohenpeissenburg (HOH), Germany	47.8 N	11.0 E	1987-2017	353	300
Mauna Loa ~~, Haiwaii~~ (MLO), Hawaii	19.8 N	155.7 W	1993 - 2017	353 / 355[2]	300
Observatoire de Haute-Provence (OHP), France	43.9 N	5.7 E	1990-2016	532	1000
Table Mountain Observatory ~~,~~ (TMO), California., US	34.5 N	117.7 W	1988-2014	353 / 355[2]	300

[1] based on data availability, [2] post 2001 upgrade.