Response to referee #1

Thank you very much for your helpful comments and for the time and effort put in to help with this study.

Abstract: Include as well differences found in the edges of the measurements (50 and 85km) and a sentence summarizing the main errors in OS temperatures.

These are now mentioned, along with systematic errors.

P9495 L6-8: some references to results from OSIRIS could illustrate this sentence

McLinden et al. (2012) has been added.

P9495: include a reference to a more detailed description of OSIRIS (maybe Murtagh et al. (2002), that you already have in your reference section but is not used in the text)

Omission of the Murtagh reference here was an oversight, reference has now been added.

P5495 L9: ’vertical spacing’ instead of ’vertical resolution’

Has been changed to “vertical field of view”.

P5495 L24: ’ from around 20km’ instead of ’from the surface’

Changed “the surface” to “lower stratosphere”.

P5495 L25: ’field of view’ instead of ’resolution’

Corrected.

P5495 L28: delete ’daytime’ (temperatures from SABER are derived from the 15um emission also at nighttime)

Deleted.

P5496 L5-6: whether ’a minimum’ instead of ’an approximately’ or ’sampling’ instead of ’resolution’

Changed to “approximately 4 km vertical field of view at the tangent point”

P5496 L16: SOFIE temperature retrievals extend up to 102 km (Stevens et al., 2012)

Corrected, and added reference.

P5496 L16: ’field of view’ instead of ’resolution’

Corrected.

P5496 L17: Perhaps more precise to write ’66 and 83deg’ (Stevens et al., 2002 or Marshall et al., 2011). In fact, you write ’66deg’ in section 3.3.

Corrected.

P5496 L18: include the version number (v1.2) of SOFIE temperatures you compare with

Added version number.

P5496 L23: What is PCL vertical resolution and sampling?

The vertical resolution of 1008 m is now stated.

P5497 L8: ’lower by 2-3K’ (see Remsberg et al., 2008)

Corrected

P5497 L14: Include reference to radiance at 278nm in Fig.1a.

This does not seem like an appropriate place to reference Fig1a. The 278 nm radiance is now further discussed at P5499 L11.

P5498 L28ff: Fig1b suggests that temperatures from 318.5nm match the MSIS profile around 65km, better than those from 347.5nm (although I understand that in the retrieval at 347nm you use T_o and n_o taken at 72km and not at 85km as in the plot). Could you justify the choice of the initial altitude of the 347nm retrievals at 72km instead of, let’s say, 60-65km?

The justification of altitudes is now discussed at P5499 L11.

P5498 L3ff: On the one hand and according to Table 1, the systematic error due to T_o decreases one half in 5km and the noise error at 84km is smaller than 2K. On the other hand, the Tk error from the O2 A-band measurements is significantly smaller at 90km than at your z_o=85km (P5498 L2). I would not be surprised to see smaller errors in the 80-85km altitude range if the initial altitude of the retrieval were 90km. Have you done any test in that sense? Would that reduce the differences (in particular, with PCL) in the upper mesosphere?

It is true that the A-band T error decreases with increasing altitude at 85 km, however the error in the OSIRIS signal is also increasing greatly with altitude. 85 km was (empirically) found to be the optimal height to balance these uncertainties to get the most appropriate temperatures. This is now discussed in the text.

P5499 L8: Specify what ’corresponding MSIS [M] profile’ is

Has been replaced with "co-located MSIS background density profile"

P5500 L20: Given that the mesospheric temperature presents typical scenarios (polar summer and winter, mid-latitudes, etc...), comment on any significant latitudinal or seasonal dependence of the estimated errors (I think you suggest there is some when attributing differences to larger errors in the southern high latitudes P5506 L15).

We did not find any significant seasonal dependence in the estimated error, but a slight latitudinal dependence was observed, SH mid latitude uncertainties are ~1.5 K higher than the mean uncertainty near 75 km, now mentioned in the text.

P5501 L5ff: You do not mention temperature errors due to uncertainties in other important parameters, as n_o (which I guess comes from MSIS), n(z) (which,should have errors coming from, e.g., cross section uncertainties), not accurate inclusion of multiple scattering (you mentioned that has a large impact in the comparisons later on, e.g., with SOFIE), OSIRIS pointing; O3 and NO2 absorption. Include estimations of systematic errors due to these sources in your error budget or, at least, comment on them explicitly in the text

Uncertainty due to pointing error was mistakenly overlooked, and is now accounted for and discussed.

Without a sophisticated height and albedo dependent scattering model it is not possible to determine the uncertainty due to neglecting multiple scatter within the retrieval. In this section, we now discuss that there is the potential for uncertainty in the retrievals at low altitudes due to the omission of multiple scatter, especially in regions where the albedo is relatively high. Although it is not included in the error table, the caption now notes that "total systematic" does not take this effect into account.

Relative n profiles (including n_o) are retrieved from OSIRIS radiance observations. Errors in the density retrieval come from instrument calibration, instrument noise, solar flux uncertainty (all already discussed), pointing errors (now discussed), and cross section uncertainty (now mentioned as insignificant, i.e. < 0.09 K at all altitudes).

As discussed in the text, the altitudes regimes were chosen specifically to avoid regions where O3 and NO2 absorption is significant, i.e. NO2 and O3 absorption is insignificant above 45 km at 347 nm, and O3 absorption is insignificant above 72 km at 318 nm.

P5501 L15: Indicate the criterium used for PMC detection and mention if there could be Tk errors coming from failed detections

This is now discussed, "In order to detect a PMC within a given OSIRIS scan, a predicted single-scatter radiance profile between 50 and 90 km was determined by fitting the OSIRIS 348 nm radiances outside this altitude range to a 5^{th} order polynomial. If at any altitude level the OSIRIS 348 nm radiance value exceeded 1.35 times the predicted value, a PMC was determined to be present. The fitting parameters and detection threshold were specifically chosen so that it would be much more likely to falsely predict the detection of a PMC than to fail to detect a PMC."

P5501 L25: Indicate the percentage of profiles that were filtered out. Is there any ’preferred’ altitude of the outliers where the temperature more often departs from 3.5xMAD?

Now state that less than 3% of profiles were rejected, and the altitude dependence is discussed.

P5502 L8: Delete ’is’

Deleted

P5503 L4: You could comment on the differences in temperature amplitudes (>10K) and their causes (different vertical resolutions, larger SABER non-LTE errors in inversion layers, co-location mismatch)

The max difference (just slightly less than 10 K) is now discussed, with non-common volume as a possible cause. (vertical resolutions are fairly similar, and it may not be fair to attribute the cause to SABER)

P5503 L15: The differences should be compared with the SABER and OSIRIS combined systematic errors. I think they will lie within that combined error and, therefore, can be expected and explained

Now mentions that the observed bias is within the combined systematic uncertainties.

P5503 L15: There is no discussion in terms of known biases from SABER. For example, the 2K high bias in the mid-mesosphere found here agrees with a reported SABER 2K low bias (Remsberg et al., 2008)

Now mentions that the observed bias agrees well with the previously mentioned SABER 2 K bias.

P5503 L20: You are suggesting here that a higher surface albedo could account for up to 4K error in the stratopause. That implies that you must include that source of error in your budget (Table 1), even if you have to specify that it takes place only at high latitudes

Multiple-scattering error is now discussed when discussing the error budget.

P5503: Even if the number of co-located pairs is not large, please, give some information about the behaviour of the SABER and OSIRIS differences for PM temperatures

At the end of section 3.1, “For similar comparisons between coincident PM temperature profiles, not shown, OSIRIS temperatures are typically colder than those of SABER, within 5 K near 55 km and within 3 K near 82 km. Near 75 km, OSIRIS temperatures are 0-1 K warmer than those of SABER. Although these biases are ~2-5 K colder than those using the OSIRIS AM temperatures, they are within the combined systematic uncertainties between 60 and 83 km. There was no significant AM/PM dependence found in the OSIRIS systematic uncertainties at the lower altitudes, therefore knowledge of the dependence of SABER systematic uncertainties on diurnal variation is needed in order to determine potential sources of this disagreement."

P5504 L2: ...but I can only read in Sheese et al. (2011) that 'An investigation into the source of these biases is needed', refering to the OSIRIS cold bias wrt SABER in the mesopause. Could you add a short line in the text outlining possible reasons for SABER and OSIRIS disagreement at 85km?

Sorry, I thought we had outlined reasons in the Sheese et al. paper. Lack of knowledge of O2 density and A-band background signal in the OSIRIS retrieval and uncertainties in O for SABER are now listed.

P5504 L11: I think that dividing the comparisons in latitudes but not in seasons might not be advisable because ACE-FTS biases in the upper mesosphere might be dependent on seasons (positive in the polar summer and negative in the polar winters, see Sica et al. 2008). Examine if that also happens in comparisons with OSIRIS.

Sica et al. 2008 (if you are referring to the ACE validation paper) does not mention a seasonal dependent bias in the temperature comparisons. In fact, when comparing with SABER, it explicitly states that the bias above 50 km is independent of season (and latitude). However, due to the Odin orbital geometry, OSIRIS never observes daytime conditions in the polar winter, and therefore this wouldn't be an issue.

P5504 L12: You should use ACE-FTS averaging kernels to smooth OSIRIS profiles. If not available, why using a running mean and not a gaussian, which, in principle, should be more representative of ACE-FTS FOV? Does using a gaussian change the comparisons?

ACE-FTS does not have averaging kernels as its retrievals use non-linear global least squares fitting. We have now tested this and now mention that using a Gaussian does not significantly affect the comparisons.

P5504 L17: That contradicts your finding of an improvement of 0-2K in comparisons with v3.0 with respect to v2.2. Remove 'little' and write a number.

Replaced "little" with "a less than 3 K"

P5504 L22: Add 'or Fig. 7a between 60 and 70km'

Added

P5505 L9: Delete 'using the stricter coincidence criteria'. Both criteria lead that result.

Deleted

P5505 L10: ACE-OS using the -4K-line as reference and SABER-OS behave similarly. That reinforces the results of the comparisons with SABER. It also justifies the use of the -4K-line in the plots, which you do not currently mention in your discussion. Include a comment on this.

The ~4K bias is now mentioned here.

P5505 L16: Could you include possible reasons for the significantly different behavior in the southern high latitudes above 75km?

There does appear to be slightly worse agreement above 75 km in the southern high latitudes, however it's difficult to say if this is a robust result when only comparing 10 (6 for v3) profiles. This is now mentioned in the text.

P5505 L25: Again, I would avoid grouping polar summer with polar winters, given the large difference in temperatures in these two scenarios. What happens if your sub-plots are done for different seasons instead of hemispheres?

These are all polar summer, as OSIRIS does not retrieve polar winter temperatures. This is now re-emphasised in the text.

P5506 L8: I miss discussion regarding previously reported biases of the other instruments. Specify the findings of Stevens et al. that you refer to: for example, comparisons with SABER and ACE-FTS point to a SOFIE 2-4K cold bias at 75km in the polar summer; also, Stevens et al. show a SOFIE 2-3K warm bias at 50km, which might partially explain the differences you find here

These biases are now briefly mentioned.

P5506 L14: Would comparisons of OSIRIS vs SABER for the Antarctic (including all seasons) lead to a similar results, particularly, a 8K colder bias in the stratopause? That should be the case if the bias is due to inaccurate inclusion of multiple scattering. If so, why OS-ACE does not show that larger bias in the 60S-82S box?

Unfortunately, OSIRIS really only measures daytime conditions in the Antarctic during the summer. Again, this is most likely due to the fact that we are only comparing 6 profiles in this region. This is now mentioned in the text.

P5506 L17: This is another reason to include in Table 1 systematic errors coming from wrong multiple scattering.

Multiple-scattering error is now discussed in the error budget discussion.

P5507 L7: OSIRIS warming around 80km is not as large as that of PCL (difference at the peak is almost 20K and OS barely shows a positive gradient). According to PCL Tk geophysical variance, that should not be due to time mismatch. What is the distance between PCL location and OS measurement? You have not smoothed PCL profiles, which (I guess) have better vertical resolution than that of OS. Could the difference of the inversion layer be partially explained by the different vertical resolutions? Also, comment on the excellent agreement below 70km.

As now mentioned in the introduction, the vertical resolution of the PCL profiles used in this study is 1-km. Therefore the most likely explanation is the mismatch in location, and low OSIRIS bias, which is now discussed in the text along with the good agreement below 70 km.

P5507 L11: Why haven't you used OS averaging kernels to smooth PCL profiles to match the vertical resolutions?

As now mentioned in the introduction, the vertical resolution of the PCL profiles used in this study is 1-km. This is on the same order as the OSIRIS vertical resolution.

P5507 L22ff: The standard deviation gives information not only about the coincidence mismatch but also about the precision

Agreed, however since the natural variability within the coincidence criteria can be significant, we haven't attempted to use the standard deviations to estimate the OSIRIS precision.

P5508: Include anywhere in the summary that, except for ACE-FTS (which has a 4K bias) and at southern high latitudes, OS measures 2-5K WARMER temperatures from 60 to 80km.

OSIRIS has a cold bias in the northern high latitudes between 55 and 70 km compared to SOFIE, compared with PCL the OSIRIS warm bias is on the order of 0-2 K, and as previously mentioned SABER has previously been shown to have a ~2 K cold bias. Therefore, this statement may not be representative of the results.

P5508: Also, once the combined systematic errors are considered in the manuscript, mention whether the differences are within the combined errors or not

Discussion of differences typically being within the combined errors is now included.

FIGURES

Fig1b: What do you mean with 'Normal retrieval' in the legend? 'derived from MSIS' or even 'MSIS Tk'? Please, clarify here or in the text

Normal retrieval refers to switching from 318 nm to 348 nm at 72 km. This is now clarified in the text.

Fig. 4: You could plot SABER and OSIRIS combined error instead of the standard deviation of the difference. Also, it is interesting that restricting the coincidence criteria, the results remain the same but, since both Fig. 4a and 4b are similar, it is enough with Fig.4a and a comment in the text. Please, remove Fig. 4b.

Figure 4b has been removed, the results of which are still discussed in the text. Now show combined systematic uncertainty.

Fig. 5: What happened to latitudes higher than 72deg? Comparisons with SABER at very high latitudes would be a nice test to see if comparisons with SOFIE make sense.

There were no coincident profiles. This is due to the fact that we're only observing the highest latitudes in the summer, where there are almost constantly PMCs.

Fig. 6: What happened to latitudes higher than 60S?

Due to the Odin orbital geometry, there are no daytime southern high latitude observations at this time of year.

Figs 5 and 6: For consistency, please, use the coincidence criteria used for Fig 4a (1000km;1h) in these plots

Now using 200 km, 1 h for Figure 4.

Fig. 8: Why do you also show ACE-FTS 4K bias in v3.0 if those temperatures are not yet validated nor such a bias in v3.0 has been reported? Your comparisons show that ACE v3.0 temperatures are 0-2K colder than v2.0, at least below 80km.

Keeping the 4K line may make observing the differences between v2.2 and v3.0 easier than if it were removed.

Fig. 8: Given the small effect of tightening the coincidence criteria, please, remove right panels and leave the discussion only in the text. I would also merge the plots showing v2.2 and v3.0 together for an easier comparison of these two versions. Additionally, as in the case of SABER comparisons, plot the combined systematic error instead of the standard deviation

The plots for tighter criteria have been removed. We have kept v2.2 and v3.0 separate as putting them together looked too crowded with the standard deviations. Standard deviations have been kept and we have not added combined systematic errors, as ACE-FTS does not have estimates for their systematic error. The ACE-FTS data products do not contain estimates of systematic errors, and therefore there is no discussion of ACE-FTS systematic errors in the Sica et al. 2008 ACE temperature validation paper. The 4 K bias will have to suffice.

Fig. 11: Remove standard deviations and add combined systematic error

Both are now shown.

Fig. 13: Again, given the similarity of the three plots, I would prefer having only one of them and a short discussion in the text. Also, remove standard deviation and plot combined errors.

Now just showing the first plot, and showing both standard deviations and combined errors.

REFERENCES

Bucholtz 1995: Referenced in the text but not included in the list

Reference has been added to the list.

Gattinger et al. 2008: the doi does not correspond to the reference. Maybe 10.1139/p08-015?

10.1139/P08-015 was the doi given in the manuscript and is the doi published by CJP.

Murtagh et al. 2002: this reference is not used in the paper

Reference has been added to the text.

Stevens et al. 2012: the paper is now published

Reference has been added/corrected