

## Author response to the referee and short comments:

We thank the anonymous referees and K. P. Shine et al. for their very sound, constructive and helpful comments which helped us to significantly improve our manuscript. In the following, we provide point-to-point replies to all comments made by the referees. All page and line numbers quoted in this reply refer to the initial version of the manuscript.

### Anonymous Referee #1

#### Comments :

In the Introduction, **page 1 line 31**, it would be useful to add a brief comment about why accurate laboratory studies cannot be performed at atmospheric temperatures.

*The following comment was added to the manuscript to clarify the current situation of laboratory studies (page 1, line 31). The wording of the revised manuscript furthermore represents the fact that, in principle, CDRS and related techniques are suitable to yield laboratory results under atmospheric temperatures in the future:*

*“Due to the decrease of water vapor saturation pressure with decreasing temperature and given the sensitivity of currently available experimental setups, such studies have been performed at least at room temperature or even heated to be able to measure the weak continuum absorption. Therefore, accurate laboratory studies have not yet been performed at atmospheric temperatures. However, extrapolation of results to lower temperature leads to significant errors because the temperature dependence of continuum absorption is in general not well modeled (e.g. Paynter and Ramaswamy, 2011).”*

On **page 2, line 11-13**, it would be helpful for the reader if you were to clarify explicitly why installing high temperature blackbody sources is drastically more challenging in remote or polar observatories. These two items are key motivators of the technique described in the paper. I feel the reader's understanding of the work's usefulness would benefit from brief comments explaining the reasons this is needed.

*As noted rightly by the referee, this is a key issue for the usefulness of the proposed method which needs further clarification. Note that the calibration approach of Gardiner et al. (2012) that relies on a very high temperature source already includes a solution to this problem via a portable source that allows transferring the calibration to field measurements. To avoid misleading representation of the work of Gardiner et al. (2012), a discussion of this solution was also included in the revised manuscript. We therefore changed the manuscript as follows (page 2, line 11):*

*“Gardiner et al. (2012) implemented a calibration method based on spectral radiance measurements of a very high temperature (3000 K) blackbody source. This method is traceable to a primary standard cryogenic radiometer, and a calibration transfer for field measurements was implemented via a portable calibration source (NPL Transfer Standard Absolute Radiance Source, TSARS). This transfer of calibration for field measurements is of crucial importance because radiative closure experiments are typically carried out at remote (mountain or polar) observatories because of the low atmospheric humidity required. However, the installation of a very high temperature calibration source is highly challenging at such sites for several reasons: Many remote observatories, including the Zugspitze site, lack sufficient laboratory space with stable ambient conditions (especially temperature) for*

*the installation of a very high temperature blackbody. Accessibility of the site with heavy instruments may be a further restriction, as is the case for the Zugspitze observatory, where access is only possible by cable car. The calibration method proposed in this study offers an alternative approach to this issue and does not require access to a very high temperature calibration source.”*

On **page 4, line 5**, what are the "short time intervals"? An order-of-magnitude timescale of an hour ? (on page 4 line 28 you note that the expected IWV variability is 1 mm during 1-2 hours. I assume this is how you've decided on the magnitude intended by short time interval.

*The following text was added to the manuscript (page 4, line 5): “We thereby limited the duration of Langley measurements to less than 2 h, which, based on the results of Vogelmann et al. (2015), leads to an IWV variability of about 1 mm during the measurements.”*

*In order to avoid unnecessary repetition, page 4, line 28 was replaced by:*

*“The IWV influence was estimated based on the expected variability of about 1mm during the 1-2 h Langley measurements according to Vogelmann et al. (2015).”*

**Page 4, line 11**, says the FTIR measurements selected for the Langley plot were averaged over 4 scans. Earlier, page 3 line 4, you said the Zugspitze FTIR averages over 4 to 8 scans. Were only 4 scan measurements an explicit selection criterion for consistency or was this coincidental?

*The following statement was added to the manuscript to clarify the selection criterion (Page 4, line 11):*

*“Only spectra averaged over 4 scans were used for the Langley fit to reduce air mass and atmospheric state variation during the spectral averaging time interval.”*

**Page 10, line 4:** Some FTIR instruments, such as AERIs, have two blackbodies for calibration. I was surprised to not see a mention of this. Would there be any additional value in adding a second blackbody?

*The following discussion about possible advantages of a second calibration blackbody was added to the manuscript (Page 10, line 4):*

*“As outlined above, the use of a single blackbody calibration source is suitable for solar FTIR measurements in the NIR, contrary e.g. to the AERI instrument (Knuteson et al., 2004) that achieves radiometric calibration in the FIR and MIR via the method proposed by Revercomb et al. (1988) using two blackbody sources at different cavity temperatures. This is mainly due to the negligible influence of thermal emission by the instrument on the measured radiance in the NIR (see Sect. 4.3). Non-linear detector response represents a further issue that would require the use of multiple calibration sources. Eventual detector non-linearity can be detected in the measured spectra as spurious radiance exceeding the measurement noise in saturated regions, i.e. within saturated spectral lines or in spectral regions beyond the detector's measurement range. However, using this method, no significant non-linearity was found for the InSb detector setup used in this study. An extension of the proposed technique using an additional blackbody source at a different temperature is therefore useful when applying radiometric calibration to spectra in the wavenumber range below 2500 cm<sup>-1</sup> or when using different detectors prone to significant non-linearity.”*

## Anonymous Referee #2

### Comments:

**Page 5:** Reference is made to the use of the Kurucz exo atmospheric spectrum but is followed by the statement “Furthermore, we use the ESS proposed by Thuillier et al. (2003)...” This is confusing. I think perhaps the authors mean that the Thuillier ESS is used in addition to the Kurucz version? Is it used only for its uncertainty values or does it also contribute otherwise? A statement indicating that, resolution aside, the two are broadly equivalent (assuming they are), would be useful.

*The manuscript was changed to clarify that in the analysis, the ESS of Kurucz (2005) was used, while other solar spectra represent possible alternatives.*

*Page 5, line 9: “In this study, we use the semi-empirical synthetic ESS of Kurucz (2005).”*

*Page 5, line 12: “Alternative ESS data can be used in for the Langley calibration without further modification of the calibration scheme. Suitable choices include e.g. ...”*

*The ESS of Kurucz (2003) differs significantly from alternative ESS sources such as Thuillier et al. (2003) or Menang et al. (2013). A discussion of the implications of these differences was added to the manuscript (page 7, line 9).*

**Page 6: First paragraph.** The discussion of the blackbody thermostat accounting for variations in dome temperature is worrying. The blackbody radiance varies with temperature and should be decoupled from dome temperature, especially as dome temperature could well vary significantly throughout an extended measurement period. The variation in radiance at these wavelengths may be small, but the impact should be briefly discussed.

*This paragraph of the initial manuscript was misleading; we thank the referee for pointing out this issue. As noted rightly by the referee, accurate calibration measurements require the blackbody radiance to be decoupled from dome temperature. This is generally fulfilled by our measurements, except for cases of fast temporal variations in dome temperature. Such cases were therefore discarded from further analysis. To clarify this, the following text was added to the manuscript (Page 6, line 3):*

*“The blackbody thermostat generally compensates the effect of changing air temperature inside the FTIR dome to keep the temperature inside the blackbody cavity constant and thereby avoid temporal variation of the emitted blackbody radiance. This is demonstrated in the red line in Fig. 6, which corresponds to two times the standard deviation of all normalized blackbody calibration curves. The measurements show that within a range of ambient temperatures from 263 to 273 K, the blackbody calibration results show only very little variation. However, fast temporal variations in air temperature cannot be compensated by the blackbody thermostat instantaneously and lead to short time intervals with temporally unstable blackbody radiance. Such measurements were discarded before further analysis.”*

**Page 6, line 20:** The spectra from the Langley calculations and the blackbody radiance curves are recorded on different days. Should have a statement verifying that the local water vapour concentrations for the two days were equivalent. If they weren't, a brief explanation of why that is not significant, specifically with the blackbody spectra, is in order.

*For both Langley and blackbody measurements, dry atmospheric conditions lead to improved calibration accuracy and were therefore required for the analysis. For Langley measurements, the relevant parameter is IWV, which was required to be below 5 mm. For*

blackbody measurements, the local water vapor concentration was required to be below  $1 \text{ g/m}^3$ . The manuscript was changed as follows to clarify this:

Page 4, line 34: "Dry atmospheric conditions increase the fraction of spectral intervals suitable for accurate calibration and generally reduce the Langley fit uncertainty due to lower atmospheric optical depth. Therefore, only Langley measurements with IWV < 5 mm were included in the analysis."

Page 6, line 1: "As for the Langley measurements, dry atmospheric conditions implicate more narrow spectral intervals affected by water vapor line absorption and thereby improve the blackbody calibration accuracy. Therefore, only measurements with an atmospheric water vapor density  $\rho_{\text{H}_2\text{O}} < 1 \text{ g/m}^3$  at the Zugspitze summit observatory were considered for calibration."

**Page 7, line 9:** Referring back to the earlier comment, if the Thuillier ESS uncertainty is to be used, the Thuillier ESS must first be shown to be some manner to be equivalent to the Kurucz.

A discussion of recent ESS results was added to the manuscript (Page 7, line 9), and a new section discussing the impact of ESS on calibration uncertainty (4.2 ESS uncertainty contribution) was added on page 7, line 6. See also the reply to point (2) of the short comment by Shine et al.

**Figure 7a):** Does this represent the ratio between a pair of sample spectra, or is it the mean of all ratioed spectra, or something else?

The original wording of the figure caption of Fig 7a) was imprecise- we thank the referee for highlighting this.

To clarify the content of Fig 7a), the figure caption was changed to:

"a) Ratio of the combined calibration curves determined from the 12 Dec. 2013 and 13 Dec. 2013 Langley measurements in combination with the 24 Feb. 2014 blackbody measurements (red line) and  $2\text{-}\sigma$  uncertainty estimate (grey shaded area)."

#### **Technical/typographical corrections:**

**Page 3, line 23:** ... that incides ... would be better written as "that is incident at a 90 degree angle on"...

The manuscript was changed as suggested.

**Page 8, line 21:** ... enables to detect ... should be ...enables us to detect...

The manuscript was changed as suggested.

### **Anonymous Referee #3**

#### **General comments**

Given that this paper will potentially be used as a methodology template for similar investigations by other research groups I think there needs to be additional information added. Areas where more information is required:

1/ additional background explanatory details: There is no mention why a 2000K BB external source cannot be used as in the method described in Gardiner 2012. A concise summary is required as some readers may not have an implicit understanding of the planck function as applied in the mid/near infrared region. For background/introduction completeness a description (and possibly an equation) of how the derived combined calibration curve is applied to FTS spectra to get calibrated spectra is required. The use of a transfer standard white lamp is only mentioned in the summary/conclusion section, so again for background completeness this technique should be mentioned in the introduction as it is another valid calibration method. The authors could also comment if the combined calibration technique could be (or not) also used with a transfer standard white lamp instead of a BB external source.

*The wording in the initial manuscript considering the technique by Gardiner et al. (2012) was misleading. In fact, the method by Gardiner et al. constitutes a valid alternative to the technique proposed in our manuscript. However, for the application in closure experiments, the combined method offers improved accuracy (see also our reply to point 3).*

*The calculation of calibrated spectra via the derived calibration curve is described in the revised manuscript as follows (Page 3, line 25): "The calibration procedure consists in deducing a calibration curve  $c(\nu)$ , which is finally multiplied with the measured spectra to achieve radiometric calibration."*

*Considering the use of a standard lamp, the following text was added to the manuscript. Page 2, line 10: "Possible methods include e.g. the use of standard lamps (see e.g. Schmid and Wehli, 1995)"*

*Page 6, line 9: "Standard lamps constitute an alternative to the high-temperature blackbody that can also be used in the combined calibration scheme. However, the spectral radiance of such calibration sources is typically prone to higher uncertainty than for blackbody sources."*

2/ additional technical information on FTS spectra acquisition and set up: this would assist in experiment replicability and comparison by other research groups. An extra table could added containing FTS spectra acquisition settings (Field of view, scan rate, resolution, average SNR etc..) along with details already given (resolution, detectors, beam splitter and scan averaging). Is the FTS spectra acquisition and set up common between the Langley extrapolation technique and black body measurements? If not, then what is the effect of this.

*A table providing details on the FTIR acquisition and setup was added to the manuscript as suggested (Table 1).*

*The following text was added to the manuscript (Page 5, line 40): "The settings of the FTIR spectra acquisition were similar to the Langley measurements (see Table 2)."*

3/ further discussion on the advantages and limitations of the combined technique would be helpful.

*The following discussion of advantages and limitations of the combined method was added to the manuscript (Page 9, line 36):*

*"A central advantage of the combined method is that it provides sufficiently accurate calibration for the quantification of the NIR water vapor continuum in an atmospheric radiative closure experiment. Furthermore, the combined calibration scheme can be implemented also at remote sites including the Zugspitze summit observatory and therefore represents a suitable alternative to the method by Gardiner et al. (2012). However, contrary to the method by Gardiner et al. (2012), the combined method presented in this study is not directly traceable to a primary standard and its accuracy for applications beyond closure experiments relies on an accurate knowledge of the ESS, which, as outlined above, is a topic of ongoing research. Therefore, the presented method is currently best suited for the use in closure experiments, while future more robust constraints on the NIR ESS are expected to*

*provide the foundation for accurate low-uncertainty calibration with the combined method for other applications.”*

4/ instrument stability, instrument line shape: there is little mention of instrument stability and effects of instrument line shape changes. The method and examples given assume the instrument over the time period is completely stable. There is nearly a 9 week difference between the spectra taken for the Langley extrapolation technique and the black body measurements. Any instrumental difference over this time could bias the combined calibration curve, thus the importance of instrument stability needs reinforcing and is a requirement of any absolute radiometric technique. It should be mentioned that any change in the instrument stability or instrument line shape would require the construction of another combined calibration curve and is only valid to calibrate atmospheric spectra taken in the same configuration. Validation via self-consistency (section 5.1) assumes complete instrument stability, this is shown but only over two days. Did self-consistency hold over a longer time period?

*The following discussion of instrument stability was added to the manuscript (Page 8, line 18):*

*“Note that modifications to the solar FTIR instrument such as realignment of optical elements require repetition of the calibration procedure and the calibration results are only valid during periods with no significant change of instrument characteristics. Such changes can be detected e.g. by monitoring the modulation efficiency of the FTIR or the instrumental line shape, which is achieved via routine HCl cell measurements (Hase et al., 2013). During the time interval covered by the measurements included in this study, no significant changes of instrument characteristics were detected.”*

5/ InSb detector performance: There is no discussion on the effects of detector intensity non-linearity. Is the InSb detector completely linear (a simple literature citation would suffice)? In the case of trace analysis InSb detector non-linearity is negligible. Is this also the case for absolute radiometric studies? InSb detectors are commonly cooled with Liquid nitrogen. Did the authors encounter ice forming on the detector windows? If so what was the effect in the combined calibration method and how did they correct for it.

*The following discussion of detector non-linearity was added to the manuscript (Page 10, line 3):*

*“Eventual detector non-linearity can be detected in the measured spectra as spurious radiance exceeding the measurement noise in saturated regions, i.e. within saturated spectral lines or in spectral regions beyond the detector’s measurement range. However, using this method, no significant non-linearity was found for the InSb detector setup used in this study. “*

*Considering the influence of ice formation on the detector, the following discussion was added (Page 8, line 18):*

*“Apart from modifications to the instrument which were discussed above, the accuracy of the radiometric calibration can decrease over time due to ice formation on the liquid nitrogen cooled InSb detector in case of leaks in the detector’s vacuum enclosure (see Gardiner et al., 2012). As outlined in the companion paper Part III, the additional absorption by ice formation is most pronounced in the 3000 to 3400  $\text{cm}^{-1}$ -range and was estimated using lamp spectra routinely recorded with the solar FTIR. The maximum influence of ice formation on the measured radiance was 1.6% at 3200  $\text{cm}^{-1}$  and was included in the uncertainty budget of the closure experiment.”*

## **Specific comments**

**Page 1, line 23** - The authors should define the NIR range, i.e. XXXX to XXXX  $\text{cm}^{-1}$ .

*The wavenumber range was added at page 1, line 23: "NIR, 4000 – 14000 cm<sup>-1</sup>"*

**Page 2, line 5** - Possibly remove the word 'unfortunately', it is superfluous in this context.

*"Unfortunately" was removed from the manuscript at page2, line 5.*

**Page 4, line 3** – Define 'short time scales'.

*Page 4, line 3 was changed to "on the time scale of few hours"*

**Page 4, line 7** – Mention what ray-tracing code package or algorithm was used in this study to allow experiment replicability by other groups if they choose.

*The following text was added to the manuscript (page 4, line 7): "In detail, the ray tracing routine of the PROFFIT software (Hase et al., 2004) was used for air mass calculation."*

**Page 4, line 11** – 4 scans are stipulated. 4 to 8 scans are mentioned on page 3 line 8. Can the authors clarify throughout this study the number of scans that are averaged?

*The following statement was added to the manuscript to clarify the selection criterion (Page 4, line 11):*

*"Only spectra averaged over 4 scans were used for the Langley fit to reduce air mass and atmospheric state variation during the spectral averaging time interval."*

**Page 4, line 30** – Could the authors elaborate why an air mass of 9 was chosen.

*The following text was added to the manuscript (Page 4, line 30):*

*"The air mass threshold of 9.0 was chosen because beyond this value, significant deviations from the linear relation according to Eq. 2 can be observed, which indicate inaccuracies in the air mass calculation."*

**Page 4, line 34** - Could the authors elaborate why the lower limit of 10 scans was chosen.

*The following text was added to the manuscript (Page 4, line 34):*

*"The uncertainty of the Langley fit increases with decreasing number of spectra, and the threshold for the minimum number of spectra was adjusted in order to achieve a Langley uncertainty comparable with the other contributions to the calibration uncertainty budget (see Fig. 6)."*

**Page 5, line 9** – It is unclear how the two ESS's are combined. Could more detail please be given?

*The manuscript was changed to clarify that the different ESS versions only represent possible alternatives and were not combined in the analysis.*

*Page 5, line 9-10: "In this study, we use the semi-empirical synthetic ESS of Kurucz (2005)."*

*Page 5, line 12: "Alternative ESS data can be used in for the Langley calibration without further modification of the calibration scheme. Suitable choices include e.g. ..."*

**Page 7, line 7** - Why is ESS not part of the total error budget? Assuming that in the water continuum measurement and model comparison exercise ESS uncertainty cancels implies that the ESS uncertainty is purely systematic, the random component will not cancel.

*An extensive discussion of the ESS-related uncertainty was added to the manuscript (see Sect. 4.2, added at page 7, line 6, see also the reply to the short comment by Shine et al. below)*

**Page 8, line 26** – For completeness and replicability, could the version of LBLRTM be stated.

*The version number was added to the manuscript (page 8, line 26): "...with the LBLRTM\_v12.2 radiative transfer model ..."*

**Page 9, line 31** - The term 'MIR' is introduced. Possibly not needed, or if required, then the wavenumber range should be stated.

*The wavenumber range was stated (page 1, line 35): "MIR, 667 - 4000 cm<sup>-1</sup>"*

**Page 10, line 4** - For completeness, the paragraph starting at this line should also stipulate the ESS uncertainty independently or as part of the overall uncertainty budget.

*The ESS uncertainty is mentioned in the corresponding paragraph (Page 10, line 5): "...and the accuracy of the extra-atmospheric solar spectrum."*

**Page 10, line 12** - Possibly replace the statement "NIR spectral range under atmospheric conditions", replaced with "NIR spectral range under a defined limited range of atmospheric conditions". This is to clarify to the reader that this calibration technique has been tested and is valid in a narrow range of optimal atmospheric conditions (no clouds, low water vapour content).

*The following text was added to the manuscript (page 10, line 9):*

*"Note, however, that the presented calibration method and the validity of the corresponding uncertainty estimate rely on a narrow range of atmospheric conditions, most notably the absence of clouds and low atmospheric water vapor content."*

**Figure 2** – The legend does not stipulate at what wavenumber this Langley plot is for. I.e. 7000cm<sup>-1</sup>? Also to avoid ambiguity, could the abscissa title 'air mass' be replaced with 'moist air mass'. Is this correct?

*The wavenumber range was added to the legend: "The Langley plot is based on radiance measurements in the 4300 to 4350 cm<sup>-1</sup> range". The abscissa title was changed to "water vapor air mass"*

**Figure 3** – The ordinate axis symbol  $c_{lan}(v)$  is first encountered by the readers in this figure. A definition of  $c_{lan}(v)$  should be given before figure 3 is referenced.

*The following change was made on page 5, line 18 to introduce  $c_{lan}(v)$ :*

*"Figure 3 shows the selection steps applied to the Langley calibration results  $c_{lan}(v)$ ..."*

**Figure 7, plot b** – Little information is conveyed in this plot. The authors could think about how to better display the information they want to portray to the reader

*The manuscript was changed as follows to clarify the general message of Fig.7b (page 9, line 1): “Figure 7b shows the mean measured (black) and synthetic (red) radiance for this set of spectra. It illustrates the very good general level of agreement between calibrated and synthetic spectra.”*

### **Short comment by K. P. Shine et al.**

The work presented by Reichert et al. (2016) is an important contribution to the quantitative use of ground-based sun-pointing spectroscopy. However, we are concerned by their claim that they have a superior calibration to that achieved by Gardiner et al. (2012). As detailed below, the approaches to calibration in the two papers are fundamentally different - Gardiner et al. (2010) perform a full radiometric calibration whereas Reichert et al. (2016) perform a point calibration to an assumed extraterrestrial solar spectrum (ESS). We do not believe the claim that the authors have a superior calibration to Gardiner et al. (2012) is sustainable, and we suggest that they de-emphasize this point in a revised version of their paper.

Our comments are not meant to detract from the central point that the technique adopted by the authors is a useful practical method in the field, and would become more so, if uncertainties in ESS were to decrease.

#### *General remark:*

*We agree with the short comment by Shine et al. on the fact that the calibration method proposed in our manuscript cannot be considered superior to the calibration implemented by Gardiner et al. (2012) in general. We furthermore agree that there is a fundamental difference between the two approaches that makes a general comparison inappropriate.*

*The statements made in the initial manuscript are only valid for the specific experimental setup at the Zugspitze site and the specific goal of the calibration, namely water vapor continuum quantification in a radiative closure experiment. We added a more precise discussion of these prerequisites and de-emphasized the comparison to the method of Gardiner et al. (2012) in the revised manuscript to avoid the misleading conclusion that our approach provides a superior calibration in general. The corresponding changes in the manuscript are listed in detail below the specific comments by Shine et al.*

### **Major comments**

1. Fundamentally, the Gardiner et al. (2012) calibration is a full radiometric calibration which is ultimately traceable to a primary standard cryogenic radiometer (see especially Section 2(d)(i) of their paper). By contrast, Reichert et al. (2016) perform a point calibration to an assumed extraterrestrial solar spectrum (ESS), which itself has significant uncertainties, as detailed in the next comment. It is therefore not an equivalent calibration, and the relative level of the uncertainties cannot be compared. We believe a more fundamental calibration could be possible with the authors' system in the future, if they used their blackbody unit as a calibration source (in the present paper, it is used to infill between calibration points derived using the Langley method, rather than for calibration itself).

The blackbody results offer the potential for a full radiometric assessment of the measurements but this would require an independent assessment of the manufacturer specifications and traceability to an appropriate standard.

However the consistency of the methods could be assessed by making a comparison of the Langley and blackbody data at the selected Langley points and see if they agree within the assumed uncertainties in the absolute radiometric values.

*The fundamental difference between the method of Gardiner et al (2012) and the combined calibration scheme was emphasized in the revised manuscript (Page 9, line 38):*

*“However, contrary to the method by Gardiner et al. (2012), the combined method presented in this study is not directly traceable to a primary standard and its accuracy for applications beyond closure experiments relies on an accurate knowledge of the ESS, which, as outlined above, is a topic of ongoing research.”*

*Concerning the comparison of blackbody results with the Langley calibration, the following text was added to the manuscript (Page 7, line 34):*

*“This low uncertainty contribution also shows that the shape of the calibration curves derived from blackbody and Langley measurements are in good agreement. However, a comparison of the absolute calibration relying solely on blackbody measurements with Langley results is not feasible with the Zugspitze instrumental setup. This is due to the fact that for blackbody measurements signal losses due to the optics setup bias the absolute level of the blackbody calibration curve, which, however, does not influence the accuracy of the calibration with the combined method presented in this study. ”*

2. We question the central assumption that the ESS in the near-infrared (NIR) is so robustly known that it can be used as a calibration point for ground-based measurements, following a Langley extrapolation (at least, for the purposes to which the calibration is put, in Part III).

There is a significant recent literature that casts doubt on this assumption and analyses differ by 5-10 %.

The main sets of satellite-based measurements of ESS in the NIR are from measurements using various versions of the SOLSPEC instrument dating back to 1983 (see e.g. Thuillier et al. 2015). Reichert et al. use the analysis of Thuillier et al. (2003), which is often referred to as ATLAS3. Thuillier et al. (2014) present an analysis of more recent SOLSPEC measurements (SOLAR2) which were about 7-10 % lower than ATLAS3. A subsequent paper by Thuillier et al. (2015) reanalysed the data from these measurements (producing SOLAR2rev), concluding that they are closer than SOLAR2 to ATLAS3, and indeed that ATLAS3 was more reliable. However, that analysis has been challenged by Weber (2015).

Part of the issue is that the ESS measured by SOLSPEC used in the SOLAR2 analysis drifted upwards by about 7% over the period of 2 years from its first deployment in 2008. Thuillier et al. (2015) conclude that the later measurements are more reliable, as the drift was due to decontamination, while Weber (2015) contends that the drift was due to instrument degradation. Weber (2015) summarizes other recent satellite and surface-based measurements (including Bolsee et al. (2014) and Menang et al. (2013)) that broadly support the lower value in SOLAR2. Reichert et al.'s measurements can shed no light on this disagreement, but their error analysis needs to reflect this uncertainty.

We recognise that at 9(22-23) Reichert et al. acknowledge that the closure validation “does not provide information on the accuracy of the used ESS”, but we feel that this clear statement comes too late in the paper and does not lead to the logical conclusion that this impacts significantly on the claimed accuracy of their calibration. This is especially so, as at 3(38) the ESS is characterised as “known”.

We suggest that the revised version of this paper reflects the ongoing debate about the ESS and fully acknowledges the impact of this on their claimed uncertainty.

As noted in comment 1, this is one reason why we feel the authors should be much more careful in drawing comparisons between their claimed uncertainty and that of Gardiner et al. (2012) when the calibration methods are so fundamentally different.

A comment on Part III of this series of papers will note that the ESS uncertainty significantly compromises attempts to derive the water vapour continuum absorption by comparing observed radiances (with an uncertain ESS) with modelled radiances using a specified ESS; it is not straightforward to attribute their derived radiance residuals to errors in ESS or to the effect of the continuum absorption.

*We thank for pointing out recent research that indicates that the knowledge of the ESS in the near-infrared is less robust than previously assumed. We agree with Shine et al. that this issue may affect the absolute radiometric accuracy of the calibration proposed in our manuscript. However, for the aim of the study, namely the quantification of the water vapor continuum in a closure study, inaccuracies in the ESS are only of very minor importance due to the design of our continuum quantification analysis. This important issue is outlined in our reply to the comments to the companion paper Part III. The importance of ESS inaccuracies for the calibration is discussed as follows in the revised manuscript (Page 7, line 6):*

*“A further uncertainty contribution is associated with the ESS used in the Langley calibration. While no uncertainty estimate was provided by the authors for the spectrum of Kurucz (2005), the 2- $\sigma$  uncertainty of the Thuillier et al. (2003) spectrum is reported to be in the range of 1.2 % at 4000  $\text{cm}^{-1}$  to 1.8 % at 8000  $\text{cm}^{-1}$ . Menang et al. (2013) state an uncertainty (1  $\sigma$ ) of 3.3-6.0 % for their ESS derived via the Langley method.*

*However, recent studies on the NIR ESS have yielded results which are partly inconsistent within the respective uncertainties and feature differences of up to 5-10% (see e.g. Menang et al, 2013; Bolsee et al, 2014; Thuillier et al., 2014, 2015; Weber et al. 2015). The ongoing discussion about the magnitude of the ESS in the NIR implicates that the ESS uncertainty estimates reported by recent studies may underestimate the real uncertainty. Therefore, the absolute radiometric uncertainty of the calibration scheme presented in this study remains tentative and more definite constraints require improved knowledge of the NIR ESS.*

*However, the ESS uncertainty has only very minor influence for the main aim of this study, namely the use of calibrated solar FTIR spectra in a closure experiment for quantification of the NIR water vapor continuum. This important feature results from the fact that the same ESS is used for calibration and synthetic spectra calculation in the closure experiment and is demonstrated in the companion paper Part III. Therefore, in the context of closure experiments, the relevant uncertainty budget does not include the ESS contribution and is shown in Fig. 6.”*

*Corresponding modifications have also been made in the abstract (Page 1, line 16):” The resulting uncertainty (2  $\sigma$ ) excluding the contribution due to inaccuracies of the extra-atmospheric solar spectrum (ESS) is below 1 % in window regions and up to 1.7 % within absorption bands. The overall radiometric accuracy of the calibration depends on the ESS uncertainty, on which at present no firm consensus has been reached in the NIR.”*

*A similar statement has been included in the summary (Page 9, line 36):*

*“However, the absolute radiometric accuracy of the calibration scheme presented in this study (including the ESS contribution) has to be considered tentative due to the fact that the magnitude of the NIR ESS is a topic of ongoing debate. Future ESS studies are expected to resolve this issue and improve the reliability of the calibration presented in this study for general applications beyond closure experiments.”*

3. There is significant circularity in the authors' attempt to assess their calibration using a radiative transfer model. They calibrate their ground-based measurements using an assumed ESS, and then use that same ESS in the model to simulate the surface irradiance, and so there is an interdependence between the measurements and the model. The degree of agreement then depends on the ability of the radiative transfer model to simulate the atmospheric transmittance (which is the topic of Part III of this series), which requires knowledge of both the atmospheric state and spectroscopic parameters.

Reichert et al. do not test whether this is the case, even though the slope of their Langley analysis would yield optical depth.

We suggest that the wording in Section 5.2 is altered to make clear that this is a relatively weak consistency check and that it does not, in any sense, act as a validation analysis (see e.g. 8(20), 8(33) and 9(27)); the abstract and conclusions should also be altered accordingly.

*The following discussion was added to the manuscript:*

*Page 8, line 20: “A further consistency check of the calibration error estimate provided in Sect. 4 can be obtained by a closure of calibrated spectra with synthetic solar absorption spectra obtained by radiative transfer model calculations, which enables us to detect any*

large deviations of the real calibration accuracy from the uncertainty estimate given in Sect. 4.”

Page 9, line 26: “This extensive agreement of the mean residuals with the uncertainty estimate further substantiates the calibration uncertainty budget presented in Sect. 4. However, the closure analysis relies on an accurate representation and a comprehensive uncertainty budget of the atmospheric optical depth obtained via the LBLRTM calculations. This uncertainty budget is presented in detail in Sect. 6 of the companion paper Part I. A comparison of the model results to the atmospheric OD derived directly from the Langley measurements shows very good agreement within the uncertainties as outlined in Sect. 4 of the companion paper Part III. Note that since for both Langley calibration and model calculations the same extra-atmospheric solar spectrum is used, the closure analysis does not provide information on the accuracy of the used ESS. In addition to the calibration uncertainty, further sources of radiance uncertainty contribute in the closure setup, e.g. IWV uncertainty or uncertainties related to the water vapor continuum. Therefore, the closure analysis does not enable a full validation of the calibration uncertainty budget. Instead, the analysis provides an indication that the calibration uncertainty budget excluding the ESS contribution presented in Sect. 4 contains no major underestimation of the real uncertainty.”

Page 10, line 7: ” The calibration results are substantiated by investigation of self-consistency for different calibration measurements and radiative closure with line-by-line model calculations. Both efforts indicate the validity of the 1.0–1.7 % uncertainty estimate.”

The abstract was changed accordingly (Page 1, line 17): “The calibration uncertainty estimate is substantiated by investigation of calibration self-consistency, which yields compatible results within the estimated errors for 91.1 % of the 2500 to 7800 cm<sup>-1</sup>-range. Additionally, a comparison of a set of calibrated spectra to radiative transfer model calculations, yields consistent results within the estimated errors for 97.7 % of the spectral range.”

## Further comments

(coordinate system “page number (line number)”)

1(33) “accurate laboratory studies cannot be performed at atmospheric temperatures”. It is more accurate to say that they “have not yet been performed”. In principle, the cavity ring down systems should have sufficient sensitivity to go to lower temperatures.

As suggested, the manuscript was changed to “have not yet been performed” in 1(33)

**2(10 and 12-13)** “Gardiner et al. (2012) proposed” and “the installation of ... a ... hot calibration source is highly challenging at remote ... observatories”. This misrepresents the Gardiner et al. paper. They did not “propose” but actually implemented the procedure and they used a portable calibration transfer instrument (called TSARS – see especially 2(d)(ii) of their paper) to transfer the laboratory ultra-high temperature black body calibration to the field. This comment is also relevant to the text at **9(37-39)**.

The manuscript was changed to avoid misrepresentation of the study by Gardiner et al. (2012).

Page 2, line 10: “Gardiner et al. (2012) implemented a calibration method based on spectral radiance measurements of a very high temperature (3000 K) blackbody source. This method is traceable to a primary standard cryogenic radiometer, and a calibration transfer for field measurements was implemented via a portable calibration source (NPL Transfer Standard Absolute Radiance Source, TSARS).“

Page 9, line 37: ”Furthermore, the combined calibration scheme can be implemented also at remote sites including the Zugspitze summit observatory and therefore represents a suitable alternative to the method by Gardiner et al. (2012).“

**3(15)** Since the blackbody source is not used directly for calibration, but as a method of in-filling the Langley - ESS calibration, we suggest that this heading is changed (as is the heading at 5(36)).

*The heading at 3(15) was changed to “3.2 Blackbody radiance measurements”, while the heading at 5(36) was changed to “4.2 Uncertainty from blackbody measurements”*

**3(36-37)** The Langley technique is an “extrapolation” not an “interpolation”.

*We thank for pointing out this wording mistake. The manuscript was corrected accordingly.*

**4(17)** “thin clouds” – provided clouds remain in a single - scattering limit and they are constant in time, they should be removed by the Langley technique. This may not be the case for the high zenith angles adopted here.

*The manuscript was changed to specify the cloud influence more precisely (4/17):  
“Temporal variation in thin cloud cover in the line of sight of the solar FTIR leads to variations in the measured radiance and therefore also biases the calibration result.”*

**4(36)** We are curious as to why measurements were made so close to the winter solstice, when the minimum solar zenith angle at Zugspitze will be around 70°, entailing a long Langley extrapolation to zero airmass. Perhaps there were operational reasons for this?

*The following explanation was added to the manuscript (Page 4, line 36):  
“While sufficiently dry atmospheric conditions occur at the Zugspitze site year-round (see Fig. 3 of Part I), due to instrumentation availability issues, Langley measurements fulfilling the selection criteria were only recorded on 12 Dec. 2013 and 13 Dec. 2013..”*

**5(1)** What was the field of view of the instrument?

*The FOV diameter was added to the manuscript (Page 4, line 40): “FOV diameter for Langley measurements 0.07°”*

**5(9-16)** We were not sure, ultimately, which of the two ESS were used in this work, as two are mentioned here (“we use”).

*The manuscript was changed to clarify that in the analysis, the ESS of Kurucz (2005) was used, while other solar spectra represent possible alternatives.  
Page 5, line 9: “In this study, we use the semi-empirical synthetic ESS of Kurucz (2005).”  
Page 5, line 12: “Alternative ESS data can be used in for the Langley calibration without further modification of the calibration scheme. Suitable choices include e.g. ... ”*

**5(9-16)** We presume that the ESS was corrected for the orbital eccentricity (which is especially important here as the measurement time is close to aphelion) but this is not stated.

*As noted correctly, a statement mentioning the scaling of the ESS was missing in the manuscript. The following sentence was added at page 5, line 12:  
“The ESS was scaled to account for the Earth’s orbital eccentricity.”*

**5(27)** We note that many solar lines missing in the Kurucz ESS were presented in Menang et al. (2013).

*This important advantage of the ESS by Menang et al. (2013) is mentioned in the revised manuscript (page 5, line 16):*

*“An important advantage of the ESS by Menang et al. (2013) is that it comprises numerous solar spectral lines not included in the ESS by Kurucz (2005).”*

**6(24-25)** We do not understand this uncertainty estimate by linear interpolation and perhaps this could be made clearer. If it is a simple interpolation of uncertainty (say between 2 points with 0.5% and 1% uncertainty), then this would imply that the claimed uncertainty at wavenumbers between their Langley points can be lower than at wavenumbers at which they make their Langley calibration.

*As assumed rightly by Shine et al., the uncertainty estimate between Langley points made in the manuscript results from a simple linear interpolation between the uncertainty values at Langley points. As noted rightly, the overall uncertainty can be expected to increase between Langley points. In the manuscript, this effect is described as an additional uncertainty contribution not included in the Langley uncertainty outlined above. This contribution results from the combination of Langley results with blackbody data (see Sect. 4.3 and green uncertainty contribution in Fig. 6) and leads to the expected increase of overall uncertainty between Langley points.*

*To clarify the uncertainty contributions, the manuscript was changed as follows.*

*Page 6, line 24: “In between Langley points, the uncertainty estimate is obtained by linear interpolation between the uncertainty values at the Langley points.”*

*Page 6, line 26: “Note that, in addition to the Langley contribution, further uncertainty is induced by the combination with blackbody measurements between Langley points. This additional contribution leads to an increase of overall uncertainty between Langley points and will be outlined in Sect. 4.3.”*

**6(22-26)** More details on the low value (less than 0.72%) of the error in the Langley extrapolation would be useful, especially given the high zenith angles employed here (see above – this means the extrapolation is a long one), and the fact that the ESS axis is logarithmic. As we read the text, the quoted error is the error in the individual points on the Langley plot, rather than the error in the extrapolation to zero airmass (which is what is important here), but we may be misunderstanding.

*The error quoted in the manuscript represents the error in the extrapolation to zero airmass, which, as noted rightly by Shine et al., is the relevant parameter for the uncertainty estimate. The low uncertainty value results from the fact that the Langley results are only used in windows specifically selected for low uncertainty, while blackbody data is used in between (see page 5, line 23-30 and Fig. 3 of the manuscript for the selection criteria).*

*To further clarify this situation, the following text was added to the manuscript (page 6, line 23): “This contribution is calculated as an error-weighted mean over the 2- $\sigma$  uncertainties of the Langley calibration results for all spectral points contributing to each Langley point, i.e. it represents the uncertainty of the Langley extrapolation to air mass 0.”*

*Page 6, line 26: “The low uncertainty of the Langley fit is a result of the spectral point selection outlined in Sect. 3.1.3, which restricts the use of Langley results to spectral points with low fit uncertainty, while blackbody measurements are used to constrain the calibration curve in between these points.”*

**6(27-36)** We found it difficult to follow this paragraph and how it ultimately led to the statement that the calibration uncertainty due to mirror inhomogeneity is about 0.2%. We suggest that this paragraph could be written much more explicitly.

At lines 30 – 31 the “diurnal variation of the measured signal” is mentioned, but we were not sure what this referred to. If it is the measured diurnal variation in the solar irradiance, then

isn't this part of the Langley signal? If it refers to the "outgoing laser beam" in the previous sentence, then more details are needed. Is it a simulation of the diurnal effect, or is this monitored in real time?

The paragraph was revised to improve the description of the analysis (Page 6, line 27):

*"Furthermore, the reflectivity of the solar tracker mirrors feature spatial inhomogeneity due to dirt and aging effects. Due to non-ideal alignment of optical elements of the solar tracker, the area covered by the instrument's FOV on the tracker mirrors changes over time, i.e. depending on the azimuth and elevation of the instrument's line of sight. This leads to spurious radiance variations in the Langley calibration and increases the calibration uncertainty. To obtain an estimate of this error, the position of the instrument FOV on the tracker elevation mirror for the azimuth and elevation values encountered during the Langley calibration has to be measured. This was achieved using an outgoing laser beam aligned with the instrument's optical axis, whose position on the tracker mirrors for a given azimuth and elevation is then monitored. In the spectral regions with least atmospheric absorption, the diurnal variation of the measured solar radiance is about 5%. This variation is due to a combination of several contributions: A first contribution is due to the change of atmospheric OD with airmass as visible in the Langley plot in Fig. 2. In addition, other atmospheric effects such as temporally variable clouds contribute to the observed signal. A final contribution is due to the mirror-related effect mentioned above. A conservative estimate of the FOV-related error is obtained assuming that the observed diurnal variation is solely due to mirror inhomogeneity and that mirror reflectivity drops abruptly by this amount (5%) outside the area initially covered by the FOV. Consequently, the error estimate is obtained by multiplying the 5 % reflectivity change with the fraction by which the area within the field of view has changed throughout the time interval over which measurements contributing to the Langley fit were made, which is deduced from the laser measurements. The resulting Langley calibration uncertainty due to mirror inhomogeneity is ~0.2 % (cyan line in Fig. 6)."*

The variation is interpreted as being due to spatial inhomogeneity whereas it could also be significantly affected by the angular dependence of the mirror reflectivity which we do not think is assessed here (see Gardiner et al. (2012) Section 3(c)). If this angular dependence was a significant contributor, then the authors' "conservative estimate" that converts the 5% change into a 0.2% uncertainty would have to be revisited.

*The following discussion was added to the manuscript (page 6, line 36):*

*"Note that for the Zugspitze solar FTIR, the angle of incidence on the solar tracker mirrors does not vary significantly during measurements or in between blackbody and atmospheric measurements. Therefore, the angular dependence of mirror reflectivity is not included in the calibration uncertainty estimate. However, for different instruments, this contribution may have to be taken into account (see Gardiner et al., 2012)."*

We also did not know what the "calibration time interval" (**line 35**) meant –presumably this means the total period over which measurements contributing to the Langley fit were made.

*Page 6, line 35 was changed to the more precise wording "...throughout the time interval over which measurements contributing to the Langley fit were made."*