We would like to thank referee#1 for their effort and time spent in revising our manuscript. Please find below the response (in black) to the comments (in red). We hope that our responses adequately address all the concerns of the referee.

General comments

The blue area shown in Fig. 1 (commented in page 636 line 20) is said to represent a variability in the profile VMR due to shifts in the tropopause. The explanation for this variability is not clearly stated. For example if the tropopause should drop as it does during some cases, one would expect larger total columns but the VMR values above the tropopause should remain fairly similar. Tropopause altitude can for instance be monitored from potential temperature time-series. The variability shown in blue is therefore more likely related to seasonal changes in the UTLS which includes the photochemical processes and OH availability.

Yes we agree with the referee#1.
What we show in Fig. 1 is more related to large scale-signals or seasonal changes in the whole atmosphere above the tropopause. In addition there might be small features like tropopause folds that only imply changes of CH4 in limited altitude regions.
To make clearer our statement we propose to modify legend in Fig.1 “shift of the tropopause altitude” to “large-scale shift of the tropopause altitude”. Similarly, we will change the figure caption “…UTLS variability caused by shifts in the tropopause altitude” to “…UTLS variability caused by large-scale shifts in the tropopause altitude”. In addition, we will describe in the text what we mean by large-scale shifts: seasonal cycle of the tropopause or vertical layering affected by large-scale horizontal atmospheric waves.

It is not clear in Page 641 line 26 if the HDO/H2O retrieval is done separately and the results are used afterwards for the CH4 retrievals.

The HDO/H2O interference is simultaneously fitted in the CH4 retrieval process. We will replace the sentence “The H2O and HDO interferences are considered by including dedicated spectral H2O and HDO windows…” by “The H2O and HDO interferences are considered by simultaneously fitting dedicated spectral H2O and HDO windows…”

In Page 642 line 3, by tropospheric here do you mean the boundary layer or lower Tropospheric…? The 100 m considered in Fig 1 and the error calculation (Page 644 line 4) seems very low for a boundary layer height, and is certainly not true for all sites.

Here we mean the free lower/middle troposphere as shown in Fig. 1 (red area). This is the altitude region of the regional scale lower/middle tropospheric signals/variations, and it is what we are interested in.
In addition there can be very local surface near signals and we assume that they are typically limited to the first 100m (shown as grey in Fig. 1). This is not a regional scale signal and we are not interested in it (it is an error source). In the manuscript we call it “boundary layer signal” but it might be better to call it “local surface signal”. We will make this modification throughout the manuscript.

In the filters deployed for in situ data, like when using only nighttime data, how does extraordinary events like those during biomass burning, can be discriminated? The Eureka filter of >0.5% of its stander deviation might not do depending on the duration of such an event.

Yes, we do not discriminate such events. In fact, we are interested in observing such events, since they provide information about a regional scale process (a big biomass burning event is a regional scale event). Maybe there is a confusion throughout the paper with our terminology “background signal”. What we mean are regional scale signals. This will be changed throughout the paper.
If the biomass burning event is a very small local one, the 0.5% standard deviation filter will remove most of these data. In case of the local conditions and the biomass burning event being very stable, the 0.5% filter might fail. We think that such situation is rather unlikely.

I would strongly recommend to put all of Sect. 3 (Pairing the ground-based FTIR and surface in-situ datasets) into the appendix. Although the information is important for obtaining an appropriate reference for the background CH4 concentrations and for the comparison, these criteria do not belong to the main contribution of the paper.

Here we disagree with referee#1.
Generally the GAW measurements are made at remote sites and are mostly well representative for regional scale CH4 variations. However, the GAW data can also be occasionally affected by very local small-scale CH4 signals. It is very important to identify these local small-scale signals if we want to use the GAW data as the reference for validating the remote sensing product.
However, GAW data when retrieved from the World Data Centre for Greenhouse Gases usually do not contain information on their representativeness. Therefore, appropriate filtering has to be made by the data users in discussion with the data providers (colleagues from the Global Atmosphere Watch community are co-authors of the paper). For very remote sites like the polar sites the filtering is less critical since local influences are very unlikely. However, at other sites the kind of filtering can significantly affect the result of the comparison. An example is given in our manuscript in APPENDIX A, e.g. Fig. A4. In our study we apply rather conservative filters and our results are only valid for the filters as described in Section 3. Please, be aware that this is an AMT paper, not a ACP paper. The 'How' is just as important as the 'what'.
Different lower/middle tropospheric CH4 remote sensing retrieval approaches are currently in discussion and empirical validation will become very important. In order to make such validation studies comparable a documentation of the filtering methods is important and must be an integral part of such studies. To our knowledge our study is the first that performs such a continuous empirical long-term validation for several globally distributed sites (we are only aware of one similar validation study, Wang et al., http://www.atmos-meas-tech-discuss.net/7/1457/2014/, however, there the comparison is only made for a single polar site). Therefore, our filtering methods constitute an important reference for upcoming studies and we do not think that it should be separated from the main part of the paper.

Pg 661 line 11 and Pg 664 line 15. The authors demonstrate with a solid statistical analysis that the remote sensing results, using the method described in this and a previous paper, need to be corrected by a factor of 0.98. Based on which arguments do the authors state that the 2% difference between the remote sensing and the in situ measurements comes from the spectroscopic parameters? UTLS influence on the tropospheric product is, as mentioned by the authors, a good possibility and would also represent a systematic error. The a posteriori correction is shown to improve this UTLS dependency but might not be enough to account for the stratospheric influences.

According to our estimations, systematic errors are clearly dominated by spectroscopy parameters (see table 3). It is very unlikely that another error source can cause such large systematic errors.
A systematic error in the apriori profile shape can also produce systematic errors. However, a positive 2% bias in the retrieved lower/middle tropospheric CH4 would require that the WACCM model (source for our apriori) has a tropopause altitude, that is about 200 hPa (about 6km) too low. In our opinion, such large systematic WACCM uncertainty is not realistic.
Furthermore, the different sites have slightly different sensitivities wrt an error in the tropopause altitude (see column 3 of table 3 in the manuscript). In order to get the same error for all sites, the amplitude of the WACCM uncertainty must be anti-correlated to the error sensitivities. This would be a rather rare coincidence.

Technical corrections
Abstract. This sentence does note make much sense here. “The theoretical assessments are complemented by an extensive empirical study.”

We do not understand this comment and suggest leaving the sentence as is.

Pg 636 line 2. Sentence structure: move coma to after “…places,” and put one after “…time,”.

Ok.

Pg 636 line 11. Remove coma after “means”

Ok.

Pg 638 line 12. General explanation needed: “technique, which is based on recording solar absorption spectra in the infrared region and in the detection of the absorptionsignatures of the atmospheric gases.”

The general explanation is given in the following paragraph. In order to give the explanation earlier we will remove lines 12-16.

Pg 638 line 13. Should say “In some stations, the older…”

OK.

Pg 639 line 13. “…Table 1 and their location…”

OK.

Pg 640 line 1. specify the number of layers.

Here we show a general explanation of the retrieval process and the specific number of layer depends on the altitude of the observing instrument.
For the stations that participate in the study we have between 44 and 49 levels.

Pg 640 line 18. a priori, two words, italics (throughout the manuscript).

In a review of another paper we had this comment the other way round. And we suggest having throughout the paper “apriori”.

Pg 641 line 11. Word usage “We use the HITRAN 2008 spectroscopic data base for the forward simulations, except…, for which we use…”

OK.

Pg 642 lien 15. analytical

Ok.

Pg 642 line 17. eliminate “the error”

We will replace the sentence “…where the difference between the retrieved and the real state, (^x–x) the error, is…” by “…where the error, i.e. the difference between the retrieved and the real state, (^x–x), is…”.

Pg 652 line 26, numbers can be rounded to whole numbers (also for the other stations)

Ok.

Fig. 9 Y-axis. label “Difference…”

Ok.

Pg 662 line 24. Eliminate “this”

Ok.

Pg 663 line 4. “consistency in the correlations of all…”

Ok.