Final response to referee comments on paper acp-2019-5

First of all, we would like to thank both reviewers for their critical and constructive comments, which helped to significantly improve the manuscript. The concise letter-style of the manuscript was not fully adequate for the presentation of a new field of application of satellite data and was replaced by a more detailed edition with considerably extended analysis, discussion, and conclusions, in particular with respect to the associated uncertainties. We explicitly discuss and quantify uncertainties arising from boundary layer height, plume dynamics, and smoke aerosols in the revised version.

The data set providing the boundary layer heights was replaced by the ECMWF ERA5 reanalysis, which is available at hourly resolution. The previously used ECMWF analysis is only available at time steps of 6 hours (0, 6, 12, 18 UTC). Therefore, the maximum boundary layer height at local noon close to the time of the satellite overpass (21:30 UTC = 13:30 local time) was missed leading to underestimated heights. The usage of ERA5 provides far more realistic results. As a consequence, all city scenes, even the most polluted ones, likely comply with national ambient air quality standards, which is in line with isolated ground-based air quality measurements. The largest detected boundary layer concentration anomaly within all city radii (scene near Sacramento Airport on November 10) amounts now to $5.42\,\mathrm{mgCO\,m^{-3}}$ [3.97–5.96; 1σ].

We also prepared the companion paper amt-2019-243 (Schneising et al., 2019), which describes the underlying algorithm in detail and includes error characteristics based on synthetic data, validation of the satellite data with reference data, and comparisons to the operational product.

Below we give answers and clarifications to all comments made by the referees (repeated in italics).

Anonymous Referee #1

General comments

Reviewer: *The style of the text is at the edge of what is acceptable for scientific writing. It uses emotional and judgemental wording, (non-exhaustive) list of examples: title: "devastating"; abstract: "one of the most disastrous months in Californian history", "destructive wildfires raging", "burnt to cinders"; introduction: "the town of Paradise was wiped out", "an unprecedented instance in history"; conclusion: "The analysed fires were the latest episodes of the deadliest and most destructive wildfire season the state of California has ever faced." Most of these statements can be removed without loss of any information.*

Authors: We agree with the reviewer and changed the style of the text at the passages in question but retained the description of the statistics of the California Department of Forestry and Fire Protection, e.g., that the wildfire season 2018 has been the most destructive on record with respect to burned land area, destroyed buildings, and fatalities.

Reviewer: *Further, the manuscript is very short in making reference to previous work. More references to earlier CO work of the MOPITT, SCIAMACHY, IASI, TES, AIRS teams are required.*

Authors: We extended the list of references considerably and give a more comprehensive review of earlier CO work in the revised version.

Reviewer: *What is the scientific value of the paper? The general CO detection capabilities of TROPOMI have been published before [e.g. Borsdorff et al., 2018a,b]. Air quality issues with wild-fire CO emissions are well-known. I would argue that the scientific value is the quantitative estimation of the CO burden (in units mg m^{-3}) based on daily recurrent satellite data i.e. the evaluation of TROPOMI's capabilities for dense CO-related air quality monitoring. Comparison to the CAMS model could also be an added value since it might trigger model improvement. Currently, the methodological evaluation and model comparisons are too short and too vague to serve any of these scientific purposes.*

Authors: We clarified that the main scientific value is indeed the dense daily recurrent satellite monitoring of the CO burden and extended the analysis and discussion of the methodology and the associated uncertainties to serve this interpretation. We also present the comparison to the CAMS model in more detail and discuss differences at a high resolution of $0.1° \times 0.1°$ and the potential of model improvement.

Specific comments

Reviewer: *P4, L11: Please add a discussion on errors coming from the assumptions on boundary layer height knowledge. Discuss how boundary layer height is determined. Please also add figures or tables for typical boundary layer heights. Are boundary layer heights of a few hundred meters (at midday) realistic (P7, L5)? These boundary layer heights need to be validated. If the boundary layer is so shallow, a large fraction of the fire emissions might reach above the boundary layer due to initial thermal rise.*

Authors: We now describe that ERA5 boundary layer heights are defined as the lowest height where the bulk Richardson number, which interrelates stability with vertical wind shear, reaches the critical value of 0.25. We also added a discussion of the uncertainty arising from boundary layer height including the inherent uncertainty estimate based on a 10-member 4D-Var ensemble and the temporal variation of the boundary layer height between 13:00 and 14:00 for a satellite overpass at 13:30 local time. Furthermore, we illustrate the diurnal variations of the boundary layer heights at the analysed cities and fires and compare them to IS4FIRES injection heights to evaluate if the fire emissions might reach above the boundary layer and to determine the uncertainty arising from plume dynamics. It is important to note that the hourly ERA5 boundary layer heights are considerably larger than the boundary layer heights derived in the previous version as the maximum at midday is better sampled due to the better temporal resolution.

Reviewer: *P4, L27: Please add and discuss a figure showing "the fact that the simultaneously retrieved gases, methane and water vapour, are not considerably increased compared to the pre-fire background abundances."*

Authors: In the original manuscript, it could be seen that this is true for the most polluted scene when comparing Figure 6 to Figure 7. In the revised version, we added a figure comprising maps showing the simultaneously retrieved methane abundances for all analysed days. On top of that, deviations of methane from the pre-fire background is also implemented as an alternative quality filter in the revised version because XCH_4 is far less variable than XCO in the presence of wildfires and both gases typically exhibit similar error characteristics (Schneising et al., 2019). Hence, potential issues of the XCO data, for example due to reduced near-surface sensitivity in the presence of clouds or smoke, are clearly detected in the corresponding XCH_4 data and filtered out. The figure also demonstrates that methane is not considerably increased compared to the pre-fire background abundances and that the XCO enhancement patterns are not resembled in XCH_4. Thus, it can be excluded that the detected XCO enhancement is only an artefact as a result of light path lengthening because of aerosol scattering at the particulate matter of the smoke, because such systematic errors would affect both retrieved gases similarly.

Reviewer: *P4, L30: The CAMS comparison is too short to be of scientific value. Please add a quantitative discussion (e.g. average TROPOMI on CAMS resolution and calculate departures).*

Authors: In the revised version, the CAMS near-real-time CO analysis is shown on a finer $0.1° \times 0.1°$ grid and an additional figure is added showing departures to TROPOMI, which is used to discuss the differences in more detail.

Reviewer: *P9, L24: None of the statements in the second paragraph of the conclusion are actually conclusions based on the scientific results of the paper, but rather they are author interpretation of climate change impacts.*

Authors: The revised version includes much more conclusions based on the scientific results of the paper, e.g., concerning TROPOMI's capabilities for dense air quality monitoring on a daily recurrent basis, the potential of model improvement, and the compliance with air quality standards. We highlight that the accurate determination of boundary layer concentrations depends on reliable external mixing layer height information and that the feasibility of the analysis is subject to specific favourable conditions affecting the vertical distribution of emissions in the case of fires to ensure that most of the fire emissions stay within the boundary layer and that pyro-convection or direct injection to the free troposphere is unlikely. Parts of the former second paragraph have been shifted to the introduction or removed.

Reviewer: *Figure 1: This figure could be dropped. The total column sensitivity of the solar absorption concept is standard scientific knowledge. Does the algorithm take into account that the near-ground sensitivity might be reduced due to scattering layers such as wildfire particulate plumes (or low clouds)? If not, what is the impact on the air quality derivations - does the satellite "see" the entire column?*

Authors: The sensitivity depends for example on the spectral resolution and the fitting window used and therefore potentially changes for different instruments or algorithms. Thus, quantitative details of the sensitivity are rather a feature of the instrument and algorithm than of the solar absorption concept. Therefore, it is important to show the averaging kernels. However, as the AKs are now shown in the companion algorithm paper, the figure is dropped here. The newly implemented quality filter based on deviations of simultaneously measured methane from the pre-fire background ensures near-ground sensitivity for all scenes passing the filter. As a consequence, clouds and smoke near the origin of the fires are typically filtered out. However, in sufficient distance of the seat of fire the retrieved abundances are not affected and a quantitative analysis is still possible (e.g., in the analysed major cities), even in cases where efficient scattering in the visible spectral range is indicated by extensive plumes in the VIIRS images. The difference between scattering at clouds and wildfire smoke is the different particle size distribution leading to reduced visibility but far less scattering issues at the smaller particles of smoke in the far field of the fires in the 2.3 μm spectral range, where the satellite measurements are taken. This is supported by corresponding simulations included in the error analysis to quantify the impact of scattering at smoke aerosols.

Reviewer: *Figures 6 and 7: The figures are too dense with internal information. None of the acronyms (VC, SZA, VZA, dlnI/dx, ...) and few of the terms (sun-normalized radiance) are explained, the panels are too small, some of the panels are not even discussed (Temperature fit, ...). Recommendation: Either remove the figures entirely or just show the relevant parts e.g. the CO panels.*

Authors: Among other things, these figures should have demonstrated that the simultaneously retrieved gases are not considerably increased compared to the pre-fire background abundances, even for the most polluted scenes. As this was not realised by both reviewers, the figures are dropped and the mentioned fact is made more obvious by showing and discussing maps of methane and the newly implemented quality filter.

Anonymous Referee #2

General comments

Reviewer: *The first issue is the lack of any discussion of the retrieval algorithm or the expected error characteristics. References are provided to two ESA technical reports, but these appear not to be peer-reviewed, nor publicly available. A detailed presentation of the retrieval algorithm (in the peer-reviewed literature) is essential for establishing the provenance of the TROPOMI-WFMD CO products. While the general aspects of the algorithm might be similar to published algorithms developed for SCIAMACHY, some details will certainly be instrument-specific. Such details are extremely important.*

Authors: It is true that the two technical reports does not seem to be publicly available and we agree that a detailed presentation of the retrieval algorithm is essential. Therefore, we prepared the companion paper (Schneising et al., 2019) to fulfill this need.

Reviewer: *Similarly, details of the filtering methods (based on CO fit and water vapor absorption) will also influence the scientific interpretation of the data and should therefore also be discussed fully. (For example, how were the thresholds for CO fit and water vapor absorption determined?)*

Authors: The filtering method was updated in the revised version. The standard filter is described and tested in Schneising et al. (2019). It is a machine learning approach trained globally based on cloud data from VIIRS and seems to be rather strict at least for California during the analysed time period, which is indicated by comparison with the VIIRS cloud product for days before the fire. Therefore, a new alternative quality filter for this local application based on deviations of methane from the pre-fire background is implemented and described in the revised version to get a somewhat larger amount of utilisable scenes but retaining good agreement with the VIIRS cloud product. The corresponding methane threshold (3 times the methane random error) was chosen to distinguish systematic from random deviations. As a consequence of the approach, potential issues of the XCO data, for example due to reduced near-surface sensitivity in the presence of clouds or smoke, are clearly detected in the similarly affected XCH_4 data and filtered out.

Reviewer: *The second major area of concern is the lack of any proper validation results; the only mention of validation is a reference to an unpublished technical report. The history of satellite remote sensing demonstrates clearly that satellite products can not simply be taken 'at face value'. For satellite CO products, validation should preferably exploit in-situ CO vertical profiles measured from aircraft. If that approach is not feasible for some reason, the authors could either exploit ground-based FTIR CO retrievals or other satellite CO products. (These latter methods are less optimal than in situ-based methods because of issues related to averaging kernels.) Comparisons of satellite CO products with surface measurements of CO concentration are generally inadequate for validation because of the variability of CO in the middle and upper troposphere.*

Authors: Proper validation with ground-based FTIR retrievals, which are in turn calibrated using in-situ aircraft measurements, is included in the companion paper (Schneising et al., 2019) showing that the TROPOMI/WFMD XCO data set is characterised by a random error (precision) of 5.1 ppb and a systematic error (relative accuracy) of 1.9 ppb. Thereby, averaging kernel issues are appropriately taken into account in the validation as documented in the companion paper.

Reviewer: *Beyond these two major issues, it is not clear how the TROPOMI-WFMD CO product relates to the TROPOMI-SICOR product (as developed by Borsdorff et al.). Do the two algorithms give the same results? Are there other issues which might make one product preferable over the other? Is the TROPOMI-WFMD CO product routinely generated and publicly available? These are inevitable questions that will concern potential users.*

Authors: The companion paper (Schneising et al., 2019) also includes comparisons to the operational product, which uses the SICOR algorithm concluding that both algorithms are highly correlated and show good global agreement although the algorithms differ in several respects. Thus, the scientific and operational products are predestined to be used together with other products in an ensemble approach to benefit from the large range of respective realisations of different physical aspects in the individual retrieval algorithms. This is discussed in the companion paper and is out of the scope of this paper. The TROPOMI/WFMD CO product is routinely generated and publicly available but with time delay.

Reviewer: *My advice to the authors is to strongly consider writing and submitting a validation paper (to AMT or another appropriate journal) which directly addresses these issues. Such a paper is an essential prerequisite to the quantitative use of satellite CO data. Publica-*

tion of that paper would pave the way for this paper and increase its significance.

Authors: The correponding paper addressing the raised issues is available for public review and discussion on AMTD.

Specific comments

Reviewer: *In several places, word choice could be improved to be less sensational and more scientific. For example, in the title, 'Devastating' could be 'Severe'. Similarly, the expression 'burnt to cinders' is gratuitous.*

Authors: We revised the style of the text at several passages and changed the title as suggested.

Reviewer: *p. 2, line 8. Need reference for physiological effects of CO on humans.*

Authors: We added a respective reference (Omaye, 2002).

Reviewer: *p. 2, line 15. The text in this paragraph suggests that MOPITT and IASI CO retrievals are generally insensitive to CO near the surface, but this is overly simplistic. In fact, publications document that both of these instruments can provide useful sensitivity to CO near the surface in daytime scenes over land (i.e., in conditions of high thermal contrast). For example, see "Sensitivity of MOPITT observations to carbon monoxide in the lower troposphere," JGR, 112, doi:10.1029/2007JD008929 (2007) by Deeter et al., and "IASI's sensitivity to near-surface carbon monoxide (CO): Theoretical analyses and retrievals on test cases," JQSRT, 189, doi:10.1016/j.jqsrt.2016.12.022 (2016) by Bauduin et al. MOPITT is also equipped with near-infrared channels which can boost surface-level sensitivity in some scenes.*

Authors: We rephrased the respective passage to avoid this potential misinterpretation and extended the list of references (including, e.g., Deeter et al. (2007), Bauduin et al. (2017), and Worden et al. (2010) for MOPITT's combined TIR/SWIR retrievals) to give a more comprehensive review of earlier CO work.

Reviewer: *p. 4, line 1. If scenes with low clouds are tolerated, what effect does that have on the retrieval vertical sensitivity (averaging kernels)?*

Authors: As already mentioned in the answers to the general comments, the filtering method was replaced in the revised version by a new alternative quality screening algorithn based on deviations of methane from the pre-fire background, which filters out scenes with reduced near-surface sensitivity in the presence of clouds or smoke by detecting significant underestimations in the simultaneously measured XCH_4 data. By construction, all measurements passing this quality filter are sensitive to CO near the surface because both gases, CO and CH_4, typically exhibit similar error characteristics (Schneising et al., 2019).

Reviewer: *p. 4, line 13. The assumption that all of the pyrogenic CO remains in the boundary layer is dubious. Can it be assumed that pyroconvection out of the boundary layer does not occur? After the first day or so of burning, it is likely that CO in the boundary* *layer will start venting into the free troposphere, thus affecting CO concentrations in the free troposphere throughout the region. Finally, there seems to be no consideration of the uncertainty of the boundary layer height.*

Authors: In the revised version, we expanded the analysis, discussion, and conclusions, in particular with respect to the associated uncertainties. We explicitly added an estimation of the uncertainties of the determined concentration anomalies arising from boundary layer height and plume dynamics. The boundary layer heights are also compared to IS4FIRES injection heights to evaluate if the fire emissions might reach above the boundary layer and to determine the uncertainty arising from the vertical distribution of emissions near the source. To this end, we also discuss the local ambient atmospheric conditions (moderate to severe drought), which are favourable for dry smoke plumes being trapped in the boundary layer. We also note that there is no indication for Pyrocumulus or Pyrocumulonimbus in the VIIRS true color images. In summary, it is likely that most of the CO load stays within the boundary layer and that pyro-convection or direct injection to the free troposphere is negligible during the first days of the fire. However, partial venting to the free troposhere cannot be entirely excluded and it is concluded that unknown plume dynamics remains the largest source of uncertainty in the calculation of the boundary layer CO burden caused by wildfires.

Reviewer: *p. 4, line 28. The evidence that light path lengthening (in the presence of smoke) is insignificant is not compelling. The most credible way to prove this claim would involve validation results. If the evidence for this claim is based solely on retrieved amounts of methane and water vapor, those results should be presented in the manuscript to allow the reader to judge whether in fact retrievals of those gases 'are not considerably increased compared to the pre-fire background abundances.'*

Authors: In the original manuscript, it could be seen that the statement is true for the most polluted scene when comparing Figure 6 to Figure 7. In the revised version, we show this more comprehensively by adding a figure showing maps of methane for the analysed days demonstrating that methane is not considerably increased compared to the pre-fire background abundances and that the XCO enhancement patterns are not resembled in XCH4. Thus, it can be excluded that the detected XCO enhancement is only an artefact as a result of light path lengthening because of aerosol scattering at the particulate matter of the smoke, because such systematic errors would affect both retrieved gases similarly. In the error analysis, we also included simulations to quantify the impact of scattering at smoke aerosols.

References

Bauduin, S., Clarisse, L., Theunissen, M., George, M., Hurtmans, D., Clerbaux, C., and Coheur, P.-F.: IASI's sensitivity to near-surface carbon monoxide (CO): Theoretical analyses and retrievals on test cases, J. Quant. Spectrosc. Radiat. Transfer, 189, 428–440, https://doi.org/10.1016/j.jqsrt.2016.12.022, 2017.

Deeter, M. N., Edwards, D. P., Gille, J. C., and Drummond, J. R.: Sensitivity of MOPITT observations to carbon monoxide in the lower troposphere, J. Geophys. Res., 112, https://doi.org/10.1029/2007JD008929, 2007.

Omaye, S. T.: Metabolic modulation of carbon monoxide toxicity, Toxicology, 180, 139–150, https://doi.org/10.1016/S0300-483X(02)00387-6, 2002.

Schneising, O., Buchwitz, M., Reuter, M., Bovensmann, H., Burrows, J. P., Borsdorff, T., Deutscher, N. M., Feist, D. G., Griffith, D. W. T., Hase, F., Hermans, C., Iraci, L. T., Kivi, R., Landgraf, J., Morino, I., Notholt, J., Petri, C., Pollard, D. F., Roche, S., Shiomi, K., Strong, K., Sussmann, R., Velazco, V. A., Warneke, T., and Wunch, D.: A scientific algorithm to simultaneously retrieve carbon monoxide and methane from TROPOMI onboard Sentinel-5 Precursor, Atmos. Meas. Tech. Discuss., https://doi.org/10.5194/amt-2019-243, in review, 2019.

Worden, H. M., Deeter, M. N., Edwards, D. P., Gille, J. C., Drummond, J. R., and Nédélec, P.: Observations of near-surface carbon monoxide from space using MOPITT multispectral retrievals, J. Geophys. Res., 115, D18314, https://doi.org/10.1029/2010JD014242, 2010.