Interactive comment on “Towards verifying CH_4 emissions from hard coal mines using mobile sun-viewing Fourier transform spectrometry” by Andreas Luther et al.

Dear Anonymous Referee #1,

Thank you very much for this comprehensive review. We appreciate the level of detail and your effort. All the specific, as well as technical comments are useful and help improving this work. We answered all questions and implemented your suggestions. Answers are written in italics, changes regarding the manuscript are written in blue italics.

Specific Comments

S1: p1l1 - It is unclear if this statement is globally or just for Europe, where it seems about one-third of anthropogenic methane is from coal production. If its global, please provide a reference in the introduction. In either case, give an approximate percentage.

The sentence is reworded to: Methane (CH_4) emissions from coal production amount to roughly one-third of European, anthropogenic CH_4 emissions in the atmosphere.

S2: p1l18 - 2.5 higher do you mean 2.5 times as high as or 1.5 times higher? If pre-industrial is 680 ppb, 2.5 times higher is around 2400 ppb...

Changed “2.5 higher” to: 1.5 times higher

S3: p2l16 - This sentence needs to be reworded, it currently sounds like each of the studies of Hase 2015, Frey 2015, and Chen 2016 quantified urban fluxes of CO2 and CH4. However, only Hase quantified CO2, and only Chen quantified CH4. Frey 2015 quantified instrument bias and characterized the ILS in support of Hase 2015.

The sentence is reworded to: Recently, Hase 2015 and Chen 2016 combined several of these FTS instruments into ad-hoc networks in *the vicinity of major cities to estimate urban carbon dioxide (CO_2) and CH_4 emissions, respectively.*

S4: p2l23-26 - These last 2 sentences are a jump in topic and should be removed from this paragraph. If the authors wish to include this information I suggest it be moved to a paragraph in Sect. 2.1 discussing measurement uncertainties. If the authors also wish to keep the Frey 2015 citation, it could be moved there as well.

The sentences are moved to Sect. 2.1

S5: p3l2-10 - Please split into 2 paragraphs, one with measurements you use in this study (mobile FTS, wind lidar), one with other ancillary measurements not used here (stationary FTS, aircraft in situ, and aircraft remote sensing). You could also mention anticipating use of all data in a future study.

The paragraph is split into 2 paragraphs

S6: p4l1 - Quantify fast

The EM27/SUN internal feedback loop has a few tens of milliseconds. A typical start up phase takes a few seconds, depending on the position of the mirrors relative to the sun.
"Fast" is now quantified as: ... it supports start-up and repointing within a few seconds once the solar-tracking is interrupted.

S7: p4l12 - Please be consistent with wavenumbers or wavelength. Generally Ive seen IR measurements reported as wavenumbers, and wavelength for UV-Vis. Im getting wavelength resolution as around 1.4 nm (but you should check if you go with those units).

Spectral range is changed to wavenumbers ... and operates in the spectral range 4000 *to* 11000 cm^{-1}.

S8: p4l14-16: Im confused by dwell times and observation time. Did you only collect one ten scan average measurement per stop? I thought the 15 x symbols on Fig 4a indicated you made 15 stops, and at each stop you made multiple 10-scan averages (grey points), please clarify. Is dwell time the total time spent at one stop?

The duration for one double-sided interferogram is 12 *s. At every stop we took about 10 of these observations resulting in an average total integration time of about* 120 *s per stop. Repointing of the solar tracker und manually starting the measurements took on average another* 30 *s resulting in typical dwell times of about* 2 *min* 30 *s.*
In the manuscript, "... 120 *s total integration time per observation" is changed to* 120 *s total integration time per stop.*

S9: p5l5: Is CalPy used here as well for the FFT?

Yes, we used CalPy for the FFT and for quality filtering.

S10: p5l6: Quantify the accuracy of the retrievals here. Precision is already included elsewhere.

We performed several side-by-side measurements with different EM27/SUNs and also the Karlsruhe TCCON station during the preparations for this measurement campaign. The accuracy towards TCCON is based on roughly 3 *hours of side-by-side measurements of the mobile EM27/SUN and the Karlsruhe TCCON station in spring 2017. The calibration factor of* 0.9974 *corresponds to an average deviation of* 4.7 *ppb which is corrected for. Since this work is focused on emission estimation we only consider the difference between plume and background* XCH_4 *which is independent of the total column accuracy.*
Following sentence is added to the manuscript: The accuracy amounts to roughly 4.7 *ppb based on* 3 *hours side-by-side measurements with the Karlsruhe TCCON station in spring 2017.*

S11: p5l8: More detail is needed on how you define your quality filters here. Is there some cutoff threshold? Maybe a histogram in the supplement of the DC interferogram signal would be helpful?

We only recorded spectra when the car and instrument stood still and when there was direct sunlight without any clouds. In CalPy we set the DC-threshold to 1 % *for discarding interferograms. For mobile use, e.g. on a ship, Klappenbach (2015) uses a DC-threshold*

of 5 %. During all five transects only one recorded interferogram is discarded due to exceeding the DC-fluctuation threshold.
Following sentence is added to the manuscript: We set the DC fluctuation threshold to 1 % for discarding corrupt interferograms.

S12: p5l11 - other gases what other gases can be detected? Also I think quantified would be a better word that detected here - detection by itself is usually not particularly useful.

We agree that detection is the wrong word here. Quantification, however, is misleading as the reader could think of emission quantification which is not tested. We would therefore change detection to measurements.
The sentence is changed to: The spectral bandwidth of the EM27/SUN allows for measurements of H_2O, O_2, CO_2, CH_4, CO, (CO in combination with an additional channel described by Hase et al. (2016)) and other gases, e.g. HF and HCl, if present in significant amounts (Butz et al., 2017).

S13: p5l15 - Is this the first study to scale only the lower part of the a priori profile for EM27/SUN retrievals? Also, some examples of a priori and a posteriori profiles in e.g., the supplement would be helpful.

Butz et al. (2017) scaled the lower tropospheric part of the vertical profile only as they expected the plume of the volcano Etna to be homogeneously distributed in a certain layer between 3.2 and 4.9 km. This is somewhat similar to our cases as we expect the plume to be homogeneously distributed inside the PBL. We added a CH_4 profile example figure in the supplement (Figure 1).
Following sentence is added to the manuscript: According to Butz et al. (2017) who only scaled the relevant plume layers of mount Etna in Sicily, we only scaled the relevant layers of the CH_4 plumes.

S14: p5l23 - I understand how you determined background within a single transect, but what are the other background variations you are describing here and how did you observe them?

We observed linear trends in CH_4 concentration measurements

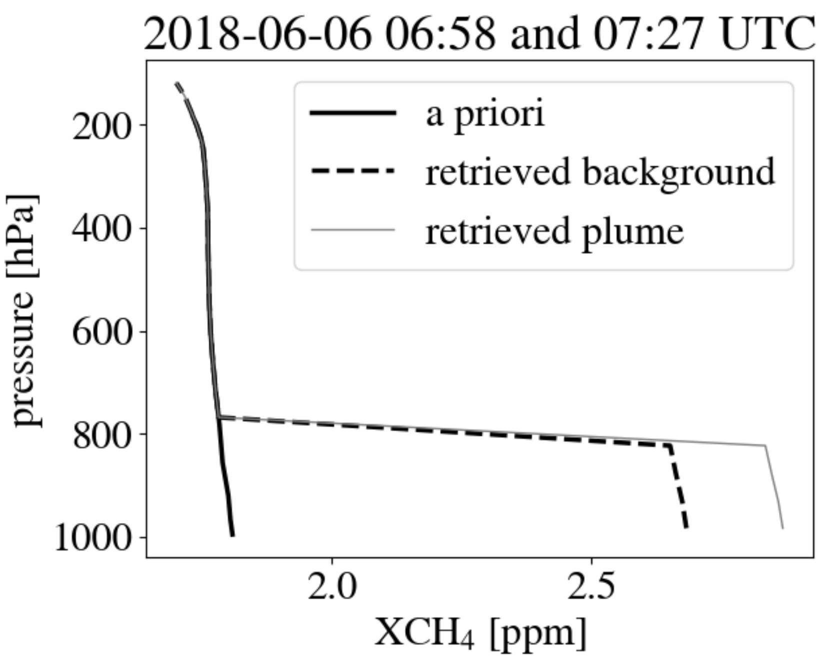


Figure 1: A-priori versus retrieved CH_4 profile. Note, that the retrieval only scales the lower part of the a-priori profile up to the the expected maximum PBL-height of roughly 800 hPa (1700 m above ground). The gray line represents an intra plume measurement.

when comparing before and after plume measurements. These trends are visible in figure 4a, 4b, 4d. We attributed these trends to the various CH_4 sources in the USCB. The stationary EM27/SUN instruments measured strong CH_4 gradients within the USCB. To avoid confusion we remove this sentence.

S15: p5l27 - Please move all information on how measurements were collected to one location (e.g., the paragraph on p4). Does this mean you were at each stop for 20 minutes typically? (If each spectrum takes 120 s).

This information is already part of section 2.1 and therefore is removed here. The next sentence is adjusted: The relative standard deviations of all measurements recorded at every stop range ...

S16: p5l30 - You have the distance between stops on p10, but it would be helpful to list the approximate distance here as well.

Information on distance between stops is added: The relative standard deviations of all measurements recorded at every stop range from 0.12 % (roughly 2 ppb) on 6 June, when most observations were taken far (> 40 km, stops every 500 m) from any source to 0.26 % (roughly 4 ppb) on 24 May, when we sampled the plume in 2 km distance, stopping approximately every 70 m.

S17: p5l32 - I agree that the 4 ppb is probably too large for a measure of instrument precision, and most likely is including real variations of XCH_4. However, it seems like it would be difficult to model these shorter term variations in XCH_4, even with the 3 wind lidars...

Yes, we agree that the 4 ppb includes atmospheric variability. Indeed, due to the turbulent nature of the plume, this variability cannot be caught by modeling. So, we decided to make it part of the error budget of the method.

S18: p5l34 - which is not possible [with measurements from] the EM27/SUN. This is not entirely true. See the abstract for B3.5 A real-time retrieval of green- house gases from portable, ground-based Fourier-Transform Spectrometers here `https://iwggms14.physics.utoronto.ca/documents/28/Abstract_Booklet_IWGGMS-14.pdf`

Thank you for bringing this to my attention. I did not know this retrieval. It needs to be tested. This part is now omitted in the manuscript: ... to discriminate background from plume enhancements of CH_4 in real-time~~, which is not possible with the EM27/SUN~~.

S19: p7l8 - Given that dyi is not infinitesimally small, it seems like it would make more sense to represent it as Delta yi

Changed to: Δy_i

S20: p7l11 - Here you convert back from an average dry-air

VMR to a column, so why do you even have Eq. 1 in the first place? It seems like it would make sense to just stick with [CH4].

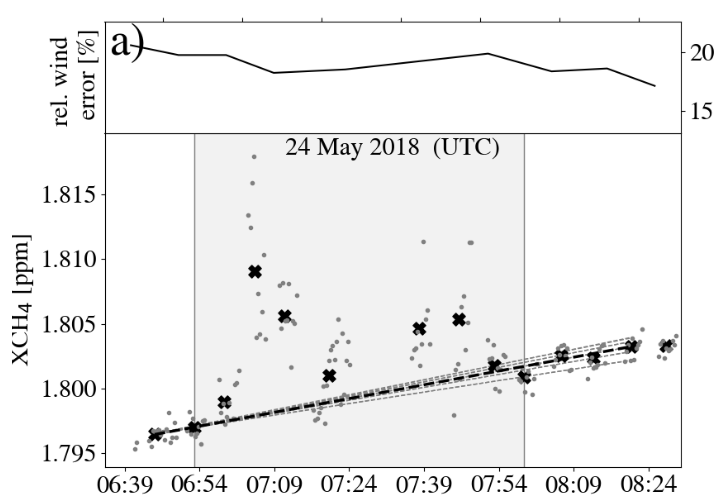
XCH_4 is required for the background removal to get ΔXCH_4 which is then used to calculate $\Delta\Omega$.

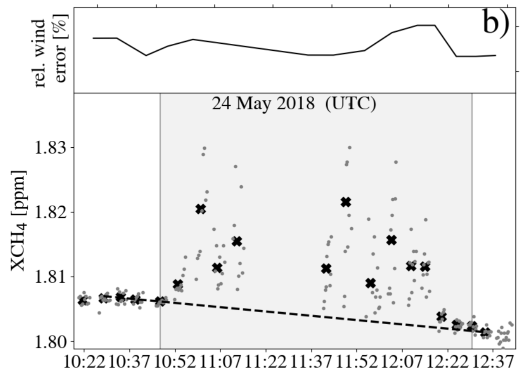
S21: p8l25 - we selected all cases - is this a subset of the 5 measurement tracks, or is this why you think the 5 tracks are good, or were there even more tracks and the 5 included here are a subset?

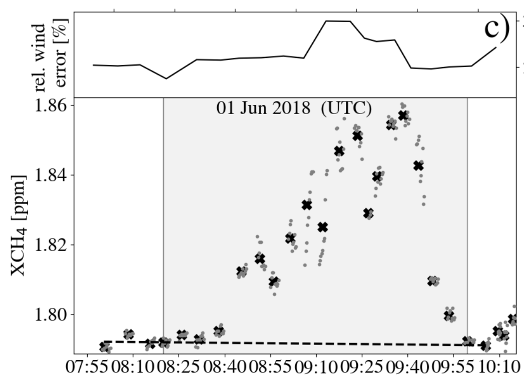
In total we performed 10 transects during the campaign period. We had to cancel two due to upcoming clouds. Three transects were finished without proper sampling of the background before and after the plume due to the lack of on the fly XCH_4 measurements.

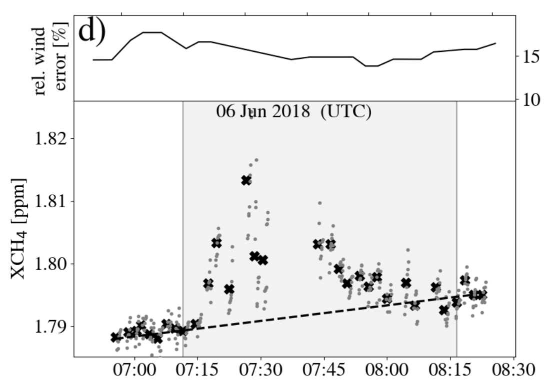
S22: p9Fig4 - I suggest you draw vertical lines separating the plume from the background. If you wanted to make the figures more information rich, you could include wind speed errors on a secondary axis.

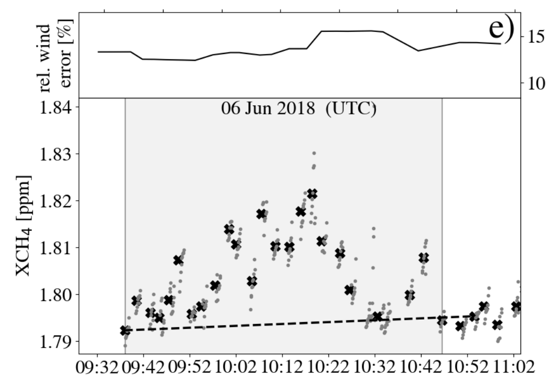
The figures are changed according to your suggestions:




b)
rel. wind error [%]
15
10
24 May 2018 (UTC)
XCH4 [ppm]
1.83
1.82
1.81
1.80
10:22 10:37 10:52 11:07 11:22 11:37 11:52 12:07 12:22 12:37


c)
rel. wind error [%]
30
20
01 Jun 2018 (UTC)
XCH4 [ppm]
1.86
1.84
1.82
1.80
07:55 08:10 08:25 08:40 08:55 09:10 09:25 09:40 09:55 10:10


d)
rel. wind error [%]
15
10
06 Jun 2018 (UTC)
XCH4 [ppm]
1.82
1.81
1.80
1.79
07:00 07:15 07:30 07:45 08:00 08:15 08:30



S23: p10l3 - It looks like 2 peaks (though if just point 6 were gone, it would look like 1 peak). Do you think you found a missing vent by chance?

We observed the easternmost of the USCB mines. There are no more known shafts upwind. The peaks are likely influenced by atmospheric variability.

S24: p11l11 - I noticed there is no discussion about averaging kernels. This is a critical omission. For total column retrievals this could lead to a 20% bias in the results. It is unclear what the effect would be here on the lower atmosphere only scaling.

This work is focused on the emission estimates and therefore we use the difference between background and plume XCH_4 *only. This difference probably depends marginally on the averaging kernel. But, we also report* XCH_4 *values which of course should be correct. With respect to S27 we performed two sensitivity studies. We retrieved the spectra again and scaled once to half of the PBL height, and once, we let the retrieval scale the full atmosphere. The mean difference between retrieved half PBL and full PBL* XCH_4 *values is* 0.6 *ppb for plume as well as background measurements. The mean difference between retrieved full atmosphere and full PBL* XCH_4 *values is* 5 *ppb. This bias is within the error budget for the FTS* *observations and therefore we omit the discussion about averaging kernels.*

S25: p11l9 - Page 10 promised a more detailed analysis here of the 6 vs. 10 discrepancy for 2 transects of a plume from the same source. However, I find this section lacking in such an analysis. Please include a greater discussion on this. Is this due to real temporal variability in emissions? This seems less likely to me. Or is this due to measurement uncertainty? This second possibility seems more likely. Maybe the uncertainty is not any of the measurements themselves, but rather from the difficulty of accounting for variability on shorter (e.g., 1 minute) timescales, and from modeling eddies, etc. I think reporting both of these numbers is very useful to the community as it shows how good reproducibility is.

The more detailed analysis is provided in section 5. The reference to section 4 is wrong and therefore adapted.

S26: p12l1 - 0.4 ppb? Compare with page 2..

Changed to: 0.3 *ppb*

S27: p13l22 - It seems unlikely for the vented methane to immediately be mixed uniformly throughout the full PBL. So I would like the authors to try another sensitivity test trying mixing to half of the PBL height (this is also somewhat arbitrary, but has been used by others (e.g., Wu et al., 2018 doi: 10.5194/gmd-11-4843-2018)). Start by re-calculating the winds for this smaller mixing depth, as well as the effect of averaging kernels (if layers are fine enough).

According to the sensitivity tests discussed in S24, a mean difference of 0.6 *ppb between half and full PBL retrievals, results in an average difference of* 0.76 % *for the estimated emissions. Averaging wind information for half of the PBL results in average differences of* 8 % *of the emission estimates. This is within the wind related error range of up to* 20 %.

S28: p14l11 - This is the first time Im learning about how you distributed emissions. This information should come in an earlier section.

Following sentence is added to the caption of Table 2: Bold numbers represent estimated emissions and errors together with the respective E-PRTR 2014 entries in the fifth column , which are the mine-wise reported values distributed evenly to every single listed shaft of each mine.

S29: p14l12 - How variable is the methane ventilation (e.g., 5%, by a factor of 2 or more)? Why is it variable?

The data we got from one individual mine suggests, that the concentration in the mine can at least double within one hour. However, we have no information on how accurate these measurements are. The methane concentration in the mine e.g. rises if a new coalbed is opened. Mining deeper (as done in parts of the USCB) also comes along with higher methane emissions as coal from deeper levels generally contains more methane.

S30: p14l13 - The personal communication reference should be omitted here since J. Swolkien is a coauthor and their contributions are listed in the Author contributions section.

Is omitted.

S31: p14l21 - A 40% difference between transects still seems large. I would like to know the mechanism for variable emissions. See also S29.

Please see S29.

S32: p15l10/p4l15 - While increasing the scan time decreased the time needed for ten scans, you should also mention that it also decreased your SNR (likely decreased by about a factor of 2).

Text on p4—15 changed to: For the last part of our deployment, we tentatively increased the sampling rate to 40kHz, which resulted in average dwell times of 60s , but also decreased the signal to noise ratio.

Text on p15—10 changed to: Enhancing the sampling frequency of the FTS decreased the dwell times significantly (only affects 6 June) but also decreased the signal to noise ratio. However, the relative standard deviation due to averaging of the FTS observations is with 2 *ppb small compared to the plume enhancements.*

S33: p15l14-15 - Of course even wind measurements onboard the truck would not provide the full picture as you are interested in the time varying winds from the source all the way to the truck. I agree though that it could be useful (e.g., these authors put a lidar in a truck: Clements et al., 2018 doi:10.1175/BAMS-D-17-0230.1)

Thank you for this interesting article. We agree that the wind history would also be important to know/model.

S34: p15l16 - Quantify the confidence here.

*Confidence added: Summarized, our approach enables the emission estimation of CH_4 with good confidence (*15 *to* 30 %*).*

S35: p15l18 - Quantify fast (1-2 hours per shaft?) and accuracy of method here.

*Added the quantities: fast = (*1 *to* 1.5 *h), accurate = (combined relative error of* 15 *to* 30 %*)*

S36: Title and p13l31 - The "towards verifying" in the title makes it sound like this is the first of several steps in regular estimates of methane fluxes. If so, what are the next steps? Providing information to policy makers or to mining companies? Adapting the method to require fewer personnel hours? Decreasing the uncertainties? Repeating measurements over a longer time? If the first part of the title were changed "Towards verifying CH4 emissions from hard coal mines" → "Quantifying methane emissions from hard coal mines in Poland" these questions become irrelevant.

Title is changed to: Quantifying CH_4 emissions from hard coal mines using mobile sun-viewing Fourier transform spectrometry

Technical comments

T1: p1l7 - distance to → from
Done!

T2: p1l7 - Move using a mass balance approach to the end of sentence
Done!

T3: p1l13 - itself → themselves
Done!

T4: p2l1 - omit however
Done!

T5: p2l5 - with → as
Done!

T6: p2l6 - With emissions of 466...
Done!

T7: p2l12 - a → a single (this emphasis lets the reader know to not look for measurements from other instruments)
Done!

T8: p2l14 - deliver → are used to measure
Done!

T9: p2l15 - omit wavelength it is implicit
Done!

T10: p2l23 - four EM27/SUN → four EM27/SUN instruments
Done!

T11: p2l27 - omit here
Done!

T12: p2l33 - used method → method used

Done!

T13: p3Fig1 - performed by → measured using
Done!

T14: p2Fig1 - EM27/SUN FTS → EM27/SUN FTS locations
Done!

T15: p3l2 - Fig. 1 → Figure 1
Done!

T16: p4l5 - omit in; distance to → of; source → sources
Done!

T17: p4l6 - Depending on the wind direction we chose the transects → We chose the transects depending on the wind direction
Done!

T18: p4l15 - our deployment, we tentatively increased → our deployment on June 6 we increased
Done!

T19: p4l16 - omit This only concerns data collected on June 6. (See also previous comment)
Done!

T20: p4l20 - the two → the standard two
Done!

T21: p4l20 - proposed by → developed by
Done!

T22: p5l3 - tracking was → tracking is
Done!

T23: p5l5 - change to For the retrieval of XCH4 from the FTS measurements we use the
Done!

T24: p5l16 - EMAC simulation results from a simulation similar to the simulation described → EMAC results from a simulation similar to the one described

I would prefer to keep it that way, since EMAC is short for "ECHAM/MESSy Atmospheric Chemistry" and then the simulation would be missing.

T25: p5l22 - omit and
Done!

T26: p5l27 - Im not sure what you are trying to say with this first sentence that is not already known. It could be safely omitted.
Done!

T27: p5l30 - in 2 km distance → within 2 km of the source.
Done!

T28: p6l4 - Three Doppler wind lidars of the type Leosphere Windcube 200S → Three Leosphere Windcube 200S Dopplar wind lidars Also, please include a refer- ence describing these wind lidars.
Done and reference added!

T29: p6l8 - in a → towards the
Done!

T30: p6l15 - The 75o → The 75o scans
Done!

T31: p6l18 - can → is
Done!

T32: p7Fig3 - EDR smaller than → EDR greater than ?
Done!

T33: p7l2 - tool for → tool typically used for
Done!

T34: p8l6 - dyi → yi and Equ. → Eq.

Done!

T35: p8l18 - Fig. → Figure
Done!

T36: p8l24 - a linear → a linear least squares
Done!

T37: p10l2 - closest other → next closest
Done!

T38: p10l4 - amounted → were
Done!

T39: p10l6 - could finish → finished
Done!

T40: p10l9 - North - I believe convention is that north and south should not be capitalized here.
Done!

T41: p10l15 - with → as
Done!

T42: p11Table2 - 1 June → 1 June, hence no E-PRTR estimate is reported
Done!

T43: p11l5 - was from → was also from
Done!

T44: p11l6 - Omit compared to the morning transects
Done!

T45: p11l5,7 - Directions here should be lower-case. Please fix capitalization throughout.
Done!

T46: p11l8 - which calculates to → from which we calculate

Done!

T47: p11l10 - error bars → errors
Done!

T48: p11l10 - of the several → of several
Done!

T49: p11l10 - along → from terms in
Done!

T50: p12l12 - once → once we
Done!

T51: p12l17 - large → larger
Done!

T52: p13l4 - CH4 → XCH4
Done!

T53: p13l16 - averaging to → averaging as
Done!

T54: p13l29 - 1 June. The latter-day → 1 June, which
Done!

T55: p14l7 - error bars → errors
Done!

T56: p14l17 - amount to → are
Done!

T57: p15l3 - Mobile FTS emission estimates are best estimated with → Our best estimate using the mobile FTS dataset is
Done!

T58: p15l5 - can be → are
Done!

T59: p15l5 - omit listed with
Done!

T60: p15l5 - Best estimated emissions amount to → Our best estimate using the mobile FTS data is
Done!

T61: p15l6 - in about → within about
Done!

T62: p15l7 - distance to → of
Done!

T63: p15l17 - a mobile → a modified mobile
Done!

Optional

These are additional comments the authors may completely ignore as they may be beyond the scope of this work.

O1: p4Fig. 2 - I am curious how the instrument is attached to the pad. Is it bolted down?

Yes, it is bolted down at three existing threads at the bottom side of the instrument.

O2: p10l6 - Im curious about what the issues were.

We also tried to use small wind-sondes to measure wind speed and direction at the location of the mobile FTS measurements (which failed due to technical issues). We interrupted the FTS measurements to set up and launch the wind-sonde.

O3: p13l30 - If you are interested in extending this paper to make it more relevant to studies without wind lidars, you could try including some other estimates of wind and comparing with fluxes derived using the more accurate lidars. E.g., using surface winds (similar to Chen et al., 2016) or STILT (similar to Wu et al., 2018).

This is planned for a future work.

O4: p15l21 - This section shows the amount of effort needed to make these measure- ments, which has implications for scaling this analysis to e.g., other coal mines and shorter revisit times. Satellite data, while supported by teams, do not require them in the field for intensive campaigns such as this one and can often cover much larger areas across the entire globe even. It would thus be interesting to know if TROPOMI could be used to calculate similar enhancements, though perhaps the footprint size is too large (7x7 km^2) and likely could only be used to get some aggregate USCB flux estimate. I agree CH4 imaging instruments would be useful including possibly GHGSat, and might even be able to be used in lieu of any ground-based mobile FTS measurements for much better scalability.

TROPOMI can detect the CH_4 outflow of the whole USCB. In combination with emission models, this can - hopefully - be compared to the stationary EM27/SUN network we also deployed in the area. We compared the stationary FTS data with the sparse TROPOMI XCH_4 measurements for the campaign period. However, the data is not mature enough for publication.