

Dear Anonymous Referee #2,

Thank you very much for this comprehensive review. We appreciate the level of detail and your effort very much. All the comments are useful and help improving this work. We answered all questions and implemented your suggestions. Your comment is repeated underneath in **bold font**, answers are written in *italics*, changes regarding the manuscript are written in *blue italics*. In some cases, you referred to specific pages, lines or equations. In the revised version these references might have changed. If applicable we have inserted the new reference in red brackets [] with respect to the revised version in change mode.

Please note that Sect. 7 “Conclusion and outlook” has been split up in the revised version into Sect. 7 “Conclusion” and Sect. 8 “Outlook”. Moreover, “Data availability” has been moved after Appendix A. New references were added in the revised version, as well as in our response here. You can find them at the end of the document.

Specific Comments:

- 1. The simulations show that daytime turbulence randomly changes the 3-D velocity field of the plume on short spatial and time scales. Is the cross-sectional method used to compute fluxes still viable under these conditions? The authors need to discuss this point, and if they feel it is, please address how the value and direction of the velocity vector is obtained for overpasses during turbulent conditions.**

*An assumption that must be fulfilled for the cross-sectional flux method is, that the tracer must have an advection component perpendicular to the overflight, i.e.  $u$  in Eq. (3). In other words, there must be a flux through the lidar curtain. In particular, Eq. (4) shows that the error in the emission rate calculation becomes very large when this wind component has small values. Varon et al. (2018) impose 2 m/s as the minimum wind speed required for the applicability of the cross-sectional flux method. This minimum value is also used by Sharan et al. (1996), explaining that above this threshold, advection dominates over diffusion. At the end of Sect. 2.1 we have therefore appended: “Varon et al. (2018) have identified 2 m/s as the minimum threshold of wind speed for the applicability of the cross-sectional flux method. This minimum value is also referred to by Sharan et al. (1996), arguing that above this threshold advection dominates over diffusion.”*

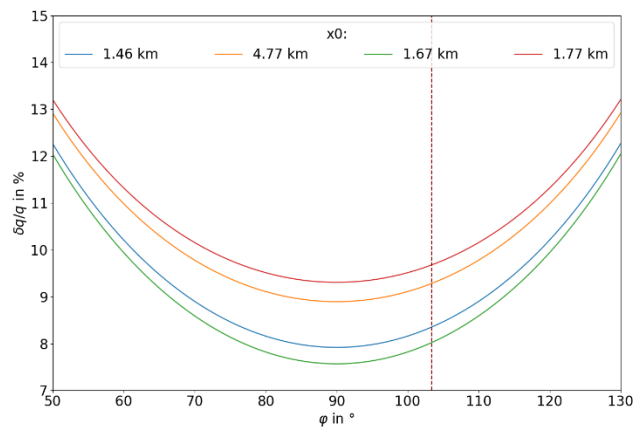
*The mean wind speed observed for the measurement is well above this threshold. Below table 1 we have inserted a remark to this regard: “The mean wind speed is well above the threshold of 2 m/s, introduced at the end of Sect. 2.1.”*

*This is also the case for the simulation. We have added the sentence “The wind speed  $u$  surpasses the threshold value of 2 m/s at all times.” to the caption of Fig. 8.*

*A description of the retrieval of the wind vector can be found in the last passage of Sect. 4. The described procedure is used under all conditions.*

- 2. It seems the precision of the lidar estimate of A depends on how many lidar measurements occur during the plume crossing. Since the laser pulse rate is fixed, it seems crossing the plume at an oblique angle (rather than at right angles) should allow more lidar measurements to be used in computing A, hence should increase the precision of A’s estimated value. Please discuss.**

*It is true that a higher number of measured data points within the plume would increase the precision of A’s estimated value. However, under consideration of the error budget estimated using Eq. (4) it becomes evident, that this would increase the overall error in the emission rate estimation. To illustrate this, we plotted the error  $\delta q/q$  over the crossing angle  $\varphi$  for the four measurement overflights, using the values given in Table 1:*



The dashed red vertical line marks the angle of HALO's flight track to the mean wind direction.

3. Several places in the paper state that the differential absorption cross section of CO<sub>2</sub> for CHARM-F is constant with altitude. For these measurements CHARM-F used the online wavelength locked to the peak of the CO<sub>2</sub> line. Due to the decreasing atmospheric pressure with altitude, the absorption cross section of the CO<sub>2</sub> line and vertical weighting functions of airborne IPDA lidar which uses this approach is not uniform with altitude. It instead peaks at the aircraft altitude. (For example, see Figure 6 in Bell et al. (2020) "Evaluation of OCO-2 XCO<sub>2</sub> variability at local and synoptic scales using lidar and in situ observations from the ACT-America campaigns." Journal of Geophysical Research: Atmospheres, <https://doi.org/10.1029/2019JD03140>.) Please address this point in the revision.

*Our measurements are based on two laser pulses, which are very well constrained in wavelength. For our purpose, we are only interested in their differences in optical depth DAOD, or more specifically, only in the DAOD enhancement induced by the plume. To clarify this, we have modified Eq. (2), thereby harmonizing it with Appendix A in Amediek et al. (2017). Note that Eq. (5) has therefore been removed.*

*It is true that the differential-absorption cross section  $\Delta\sigma$  is not constant vertically, but, in the integral of Eq. (2), we are focusing on the vertical range in which the plume provides enhanced CO<sub>2</sub> density, compared to the background concentration. The vertical extension of the plume at crossing points will be estimated by means of various atmospheric stability conditions, that can be found in the literature, as discussed in more detail in Sect. 3. It turned out that the change in  $\Delta\sigma$  over the vertical plume spread is small compared to the other error sources. So, it is regarded as legitimate to use the mean  $\Delta\sigma$ , which simplifies the calculation immensely (see Appendix A in Amediek et al. (2017)). We have replaced the term "constant" by "mean" and consistently marked  $\overline{\Delta\sigma}$  with an overscore in the entire text, to make the distinction recognizable. The fact that the mean differential-absorption cross section is only an approximation, is widely discussed in our error calculation in Sect. 3 at Eq. (8) and surrounding passages.*

*To convert DAOD to XCO<sub>2</sub>, the integrated vertical weighting function would be required. However, this is not necessary for the calculation of the emission rate. Therefore, unlike in Amediek et al. (2017), we do neither need to consider XCO<sub>2</sub> nor the vertical weighting function in this work.*

4. The simulations clearly show the transition of smooth flow of the plume during and nighttime and at low sun angles, to a more chaotic/random dispersed structure caused by turbulence at higher sun angles. This dispersal of the plume's structure seems an important

**limitation to using passive remote sensing to estimate power plant emissions. Can the authors expand on this point in the revised version?**

*We fully agree, which is why we've highlighted this aspect in the conclusion. The independence from solar irradiation is an advantage of active remote sensing to this respect. However, we are no experts for passive remote sensing and also our modeling efforts have focused on IPDA-lidar. CoMet saw the deployment of EM27 Fourier transform spectrometers, as well as the airborne passive instrument MAMAP. Synergistic studies will follow in the future; for the time being, we refer to Luther et al. (2019) and Krautwurst et al. (2021).*

**5. Please add the local sun angles to the figures of the simulated plumes.**

*The local solar altitudes  $\alpha$  have been added. We've also inserted the course of the solar altitude to Figure 8a and comment as of line 323: "In Fig. 8a it can be seen how the diurnal course of solar altitude  $\alpha$  influences the retrieved emission rates  $q$ . The random occurrence of inhomogeneities in the plume propagation, caused by local turbulence, leads to large variations in the results of successive crossings. Turbulence lags behind solar altitude because the surface needs time to heat up. It is also apparent that the emission rate deviations vary from day to day, both in intensity, as well as in dwell time.*

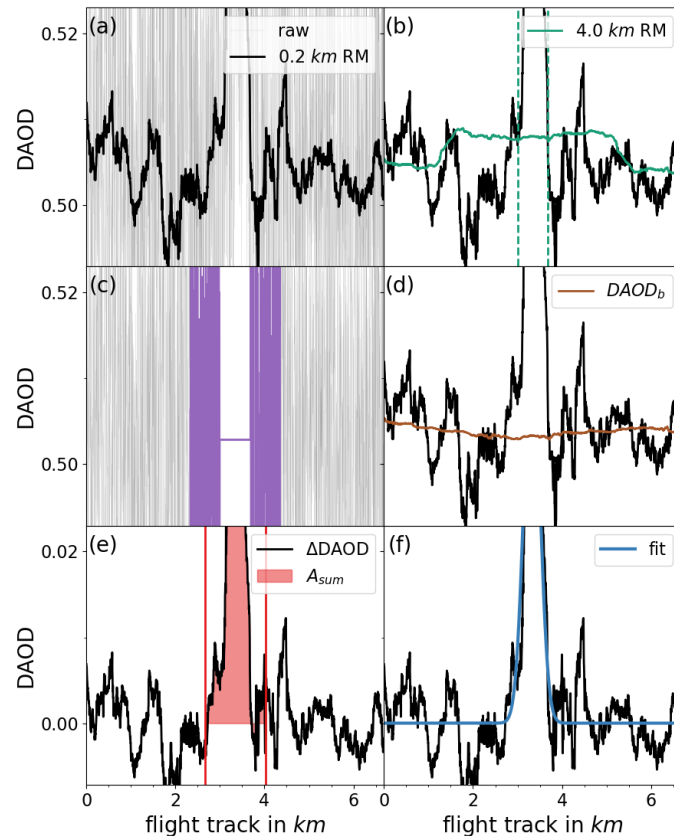
*The implications for the measurement results can be reduced by averaging over a multitude of retrieved emission rates. Next, we investigate how often the exhaust plume must be surveyed, in order to achieve a mean emission rate with satisfactory accuracy."*

**6. Equation 1 uses a mixture of laser power and laser energy, and obviously the optical power changes during the pulse. Isn't the DAOD computed from measurements of the on- and off-line laser energies (not power)?**

*In the lidar community, the term "power" is used quite laxly, although this is not accurate. In fact, it is the radiation fluxes entering the lidar telescope. We have added the reference Ehret et al. (2008), where the lidar equation for the "hard target" case is described. In this equation, the received radiation fluxes  $P_{on/off}$  are proportional to the emitted pulse energy  $E_{on/off}$  divided by an effective pulse length in the temporal domain of the lidar returns. This effective pulse length is regarded to be equal for both on- and off-line wavelengths and therefore cancels out by taking the ratio in Eq. (1). But the latter is not the case for the pulse energies, which are different for both wavelengths and therefore need to be considered in Eq. (1). More details to the shape of the lidar returns can be found in Ehret et al. (2008). For the measurement we digitally oversample the lidar returns in the temporal domain sufficiently well, which allows us to calculate both the corresponding radiation flux and the range to the targets on a shot-by-shot basis very accurately.*

**7. In general it seems that the emission plume adds CO<sub>2</sub> to an air mass that already has some variability in CO<sub>2</sub>. Hence the background also has variability in XCO<sub>2</sub>. As shown in equation 5 [now Eq. (2)], the enhancement from the plume is computed from the total DAOD minus that of the background (non-plume region). Hence variability in the background DAOD will cause variability in computed enhancement. More discussion of about the variability in the background is needed for the region measured near the plumes and on its impact on the computed enhancement from the plume.**

*It is true, that the plume adds CO<sub>2</sub> to a variable background, thereby inducing a strong local gradient in the DAOD. This plume enhancement is much more significant than variability in the background, which is why the plume can easily be located. With regard to the background, we've introduced a procedure in Sect. 2.2 that allows the background term to feature gradients on the scale of few hundreds of meters. These small-scale gradients in the background term are not nearly as pronounced as the plume gradient, or even the noise for that matter. Figure 3d shows that the background term is not constant. Here is Figure 3 again, zoomed in for better visibility:*



Evidently, our procedure does not attribute spatially smaller gradients to the background term, but incorporates them in the enhancement term. Thereby they are not distinguishable from noise and basically add an uncertainty in the estimation of the integrated enhancement  $A$ . However, we consider this circumstance to be minor compared to the actual noise. In the caption of Figure 3 we've rephrased as follows: "In (d) again a 4 km running mean over the bypassed dataset is shown in brown. This data, which has slight variability, is used as background term  $DAOD_b$ ."

Furthermore, we've added some explanatory sentences to the passage below the figure [line 179]: "At last, we execute another 4 km running mean over the raw dataset, with bypassed plumes, resulting in the background term  $DAOD_b$ , shown in brown in Fig. 3d. This procedure allows for a variability in the background term on a scale of a few hundred meters. Smaller scale gradients cannot be attributed to the background and are incorporated in the enhancement term  $\Delta DAOD$ , thereby not being distinguishable from noise."

8. In Figure 3, please inform the reader how many lidar measurements are used in the running means values, and the approximate standard deviations of DAOD for the lidar measurements with and without averaging.

In the caption we now write: "Figure 3: Plume crossing at a point source distance of 1.53 km. In (a) the gray curve shows the raw data, with a standard deviation of 5.2 %, while the black curve shows a 0.2 km (64 data points) running mean (RM), with a standard deviation reduced to 0.9 %. In (b) the green curve is a 4 km (1293 data points) RM. Green vertical dashed [...]"

9. In line 238 [now 263], it is stated the primary error in the flux estimate is from uncertainty in wind speed. This is an important point and should be emphasized in the paper's abstract and conclusion.

We agree and added to the abstract in line 17: "On average, our results roughly correspond to reported annual emission rates, with wind speed uncertainties being the major source of

*error. We observe significant variations between individual overflights, ranging up to a factor of 2.”*

*In Sect. 7 Conclusion [line 411] we write: “On the one hand, this is because the uncertainties in the wind speed are most pronounced at these times, being the major source of error in a single measurement”*

- 10. Line 238 [now 264] gives a reported value for power plant’s Co2 emission mass flow rate. Please clarify the source of the estimate. Is this some sort of average or is it based on the fossil power plant’s operating conditions at the time of the overflights?**

*It is the mean value for the year 2017, reported to the European Environment Agency. This mean value is introduced in the beginning of Sect. 3. For clarification we’ve inserted the reference. Line 254 now reads: “The reported value of 760 kg(CO2)/s (24.0 Tg(CO2)/yr) (E-PRTR, 2020) lies within the error range of [...]”*

- 11. The caption of Figure 6 states DAOD enhancements < 0.008 cannot be distinguished. It is unclear if the authors mean in the simulation, in the color plot, or in the CHARM-F lidar measurements – please clarify in the text.**

*In the caption now reads: “In a corresponding measurement, DAOD enhancement values beneath 0.008 would not be distinguishable from noise and are therefore displayed blue.”*

- 12. The LH plot in Figure 6 shows a stable linear plume extending to the edge of the plot for nighttime conditions. It would benefit the readers to know from the simulations typically how far these linear flow conditions extend in distance.**

*As a matter of fact, the linear plume is even extending the edge of the inner domain. That is why we can only report that the linear flow exceeds a distance of 17.5 km, i.e. half the extent of the domain.*

- 13. In the paragraph of line 285 [now 310], please explain in the simulation how the plume’s velocity values and directions were estimated for the turbulent conditions.**

*A description of the retrieval of the wind vector can be found right above, in the last passage of Sect. 4. The described procedure is used under all conditions.*

- 14. Line 330 [now 357] – Please be more quantitative than “mid-day” please clarify the sun angle limits or time of day limits for avoiding turbulent mixing.**

*Unfortunately, this is not possible for us. At comment 5. we have already mentioned how we observe different intensities of turbulence day by day. The specific limits in solar altitude/time of day, as well as the synoptic conditions to which they are subject, have to be evaluated by a simulation of a representative number of days in the seasonal course. As we have only performed simulations for our measurement day and the day before, this is not in the scope of this work. Nevertheless, for the two simulated days, we observe, as is to be expected, that the most pronounced turbulence occurs in the afternoon.*

*We’ve replaced the wording “midday turbulence” by the more general term “situations of high turbulence”. We feel that a more detailed engagement on this issue is better suited in the discussion Sect. 6. There, we elaborate on this matter: “At this point, we cannot derive any limits for solar altitude or local times that should be avoided, as the simulation reveals that the turbulence intensity varies from day to day (see Fig. 8 and Fig. 9). Generally, we find that the most significant turbulence occurs in the afternoon. For future campaign planning, we recommend to also perform measurements at night or in the morning, which is possible with lidar.”*

- 15. In the discussion section, please address how more accurate estimates of the wind vector in the plume could be attained in the future. For example, could co-aligned wind and IPDA lidar be used for this?**



An airborne deployment of both wind and IPDA lidar appears to be appealing. Wind lidars perform conical scans, which need up to a minute to be completed. Due to the conical scan, the lidar is pin-pointing transversal to the track of flight. Thus, the retrieved wind speed might have an additional representativeness error, depending on the inhomogeneity of the atmosphere. So, just like in the paper, the measurement result can only yield an average wind vector of the crossings vicinity and not the instantaneous velocity vector of the plume's flux through the lidar curtain. According to our DLR colleagues Benjamin Witschas and Stephan Rahm our in-house airborne wind lidar, provides data with a horizontal resolution of about 9 km and a vertical resolution of 100 m (Witschas et al., 2017). For our application it is possible to reduce the number of line-of-sights. Additionally, a new, faster scanner will be developed. All together it is realistic to achieve a horizontal resolution of less than 4 km in the near future.

Whether co-aligned wind lidar measurements are preferable to simulation data for our purpose will be subject of upcoming investigations. For instance, the MAGIC campaign will take place in 2021, during which CHARM-F will be deployed together with one of our partners' airborne wind lidar.

Concerning the CoMet campaign, three ground-based doppler wind lidar were installed in the Upper Silesian Coal Basin (USCB) and continuously performed velocity-azimuth display scans. In future studies, dedicated to the USCB, it is planned to nudge the model towards the wind lidar data, as has been done in Kostinek et al. (2020). We've appended to the first passage of the discussion: *"Future studies will examine CHARM-F measurements in the Upper Silesian Coal Basin to determine CH<sub>4</sub> emissions from coal mines. In this area, ground-based Doppler wind lidars have been installed. It is expected that nudging the simulation towards the wind soundings will result in an improvement of the wind vector estimation, ultimately reducing the overall error in the flux determination."*

Kostinek et al. (2020) was published for discussion only after our submission. We have included a citation in *line 52* of the Introduction.

**16. Several aspects of the last paragraph in the Conclusion section don't seem to apply to the main findings of the paper. Please reexamine and edit for relevance.**

Originally, we intended to put the conclusion and outlook in one Section. We now inserted a last Sect. 8 for the outlook, containing the last paragraph. Note that to do so, we've changed the last sentence *[now line 93]* in the introduction to: *"A discussion is given in Sect. 6, followed by the conclusion and outlook can be found in Sect. 7 and the outlook in Sect. 8."*

**17. Figure A2 – please include the sun angles for the plumes shown in daytime hours.**

The solar altitudes have been included and mentioned in the caption

Minor points/technical corrections:

**1. The manuscript will benefit from a check of consistency of tenses. The airborne campaign was performed in the past, while the analysis may described in the present tense.**

We have carried out the check. With a few exceptions, the tenses were consistent, however, we inserted several commas and corrected the phrasing in multiple passages.

**2. In the title, would the word "estimation" perhaps be a more appropriate word than "determination"?**

Indeed, initially we report on an emission rate estimation, with emphasis on a full error assessment. However, subsequently a considerable part of the paper is devoted to the LES simulation. This simulation is by no means used for the estimation of the power plant's emission rate, but rather seeks to provide insights for the determination of point source emission rates using cross-sectional flux method in general. Therefore, we feel that the term "determination" is better suited to encompass the content of the entire work.

3. Line 17 [now line 18] – Consider changing “we suppose.” Did not the simulations show the variability was clearly driven by turbulent mixing?

*At that point, our intention was to hypothesize based on the findings of the measurements. This hypothesis is then confirmed by the simulation. But indeed, the expression "suppose" is inapplicable. We have changed the wording, also for the subsequent sentence t: “We hypothesize that these variations are mostly driven by turbulence. This is confirmed by a high-resolution large eddy simulation that enables us to give a qualitative assessment of the influence of plume inhomogeneity on the cross-sectional flux method.”*

4. Line 30 [now line 32]: consider perhaps “assessment” instead of “stocktake”

*We would like to advocate staying with the term "global stocktake", as this is the key word that was originally used in Article 14 of the Paris Climate Agreement. (see [https://unfccc.int/sites/default/files/english\\_paris\\_agreement.pdf](https://unfccc.int/sites/default/files/english_paris_agreement.pdf))*

5. Line 43 [now line 46] – this sentence is many lines long. Please break it into logical pieces.

*The sentence has been split into: “The objective of CoMet is to investigate the fluxes of the major human-influenced GHG on local, regional, and sub-continental scales. These fluxes are to be determined more precisely than previously possible. Furthermore, supporting activities for GHG stocktaking are provided.”*

6. Page 2 and elsewhere. There are many adjectives used in the lidar description and elsewhere (e.g. on page 2: most sophisticated, small divergence, dense sequence, sophisticated, high-power etc,) whose meanings are subjective. Please either delete or replace these type of adjectives with quantitative descriptors.

*The adjectives “sophisticated” and “high-power” have been removed. In line 98 we now write: “At its core, CHARM-F consists of a pulsed, tunable laser source and a detector.” In the introduction we have changed the wording as follows: “As a result of the pulse repetition frequency (50 Hz, double pulse) and divergence (~1.5 mrad) the pattern on the ground is a sequence of overlapping footprints.”*

7. Line 60 [now 68]. The term “adaptive averaging “does not appear to be not defined or described

*High albedo yields a high signal-to-noise-ratio (SNR), while a low albedo yields a poor SNR. By applying a running mean, we can generally improve the measurement precision. However, when we measure between clouds the averaging window is limited by the size of the cloud gap. Performing a running mean that requires adapting the smooth width is what we refer to as adaptive averaging. Be that as it may, for the dataset in hand the SNR does not vary much and clouds do also not play a role. Therefore, this piece of information is irrelevant for this work, which is why we have removed the respective sentence.*

8. Line 199 [now line 222 and 223] – the distances listed in the text are slightly different than in Table 1

*At this point we are dealing with the calculation of the differential-absorption cross section  $\Delta\sigma$ . Three of the four overflights have taken place close to the point source and one further away. For the calculation of  $\Delta\sigma$  the rough distinction between these distances is sufficient, as the change in  $\Delta\sigma$  is only minor. In line 222 and 223 we have changed “=” to “≈”.*

9. Table 1- 2nd column – please check, are the crossing dimensions km or in m?

*The crossing dimension is km. “path” corresponds to the flown distance of the aircraft (i.e. the cumulative sum of the distances between successive data points). In Figure A1 the term “flight track” is used. For conformity and comprehensibility, the term “path” in the 2<sup>nd</sup> column of table 1 has been changed to “flight track”.*

**10. Table 1 – is the mean  $q$  listed for all crossings? – please clarify**

*The mean  $q$  is the average of the of the four crossings. For comprehensibility, the columns have been rearranged and a curly bracket has been inserted.*

**11. Figure 3 would be easier to interpret if the 3 boxes were labelled.**

*We assume Figure 5 showing the nested domains of the simulation was supposed to be addressed with this comment. We added labels to the domains in the Figure.*

**12. Line 365 [now 395] – please delete the word “high” and replace with the approximate accuracy attained.**

*The sentence has been changed to: “Under such conditions even a single instantaneous cross-sectional flux measurement yields an accuracy of up to ~95 %.”*

**References:**

- Amediek, A., Ehret, G., Fix, A., Wirth, M., Budenbender, C., Quatrevalet, M., Kiemle, C., and Gerbig, C.: CHARM-F-a new airborne integrated-path differential-absorption lidar for carbon dioxide and methane observations: measurement performance and quantification of strong point source emissions, *Appl Optics*, 56, 5182-5197, <https://doi.org/10.1364/AO.56.005182> 2017.
- Ehret, G., Kiemle, C., Wirth, M., Amediek, A., Fix, A., and Houweling, S.: Space-borne remote sensing of CO<sub>2</sub>, CH<sub>4</sub>, and N<sub>2</sub>O by integrated path differential absorption lidar: a sensitivity analysis, *Applied Physics B*, 90, 593-608, <https://doi.org/10.1007/s00340-007-2892-3>, 2008.
- Kostinek, J., Roiger, A., Eckl, M., Fiehn, A., Luther, A., Wildmann, N., Klausner, T., Fix, A., Knote, C., Stohl, A., and Butz, A.: Estimating Upper Silesian coal mine methane emissions from airborne in situ observations and dispersion modeling, *Atmos. Chem. Phys. Discuss.*, 2020, 1-24, <https://doi.org/10.5194/acp-2020-962>, 2020.
- Krautwurst, S., Gerilowski, K., Borchardt, J., Wildmann, N., Galkowski, M., Swolkien, J., Marshall, J., Fiehn, A., Roiger, A., Ruhtz, T., Gerbig, C., Necki, J., Burrows, J. P., Fix, A., and Bovensmann, H.: Quantification of CH<sub>4</sub> coal mining emissions in Upper Silesia by passive airborne remote sensing observations with the MAMAP instrument during CoMet, *Atmos. Chem. Phys. Discuss.*, 2021, 1-39, <https://doi.org/10.5194/acp-2020-1014>, 2021.
- Luther, A., Kleinschek, R., Scheidweiler, L., Defratyka, S., Stanisavljevic, M., Forstmaier, A., Dandocsi, A., Wolff, S., Dubravica, D., Wildmann, N., Kostinek, J., Jöckel, P., Nickl, A. L., Klausner, T., Hase, F., Frey, M., Chen, J., Dietrich, F., Nećki, J., Swolkień, J., Fix, A., Roiger, A., and Butz, A.: Quantifying CH<sub>4</sub> emissions from hard coal mines using mobile sun-viewing Fourier transform spectrometry, *Atmos. Meas. Tech.*, 12, 5217-5230, <https://doi.org/10.5194/amt-12-5217-2019>, 2019.
- Sharan, M., Yadav, A. K., Singh, M. P., Agarwal, P., and Nigam, S.: A mathematical model for the dispersion of air pollutants in low wind conditions, *Atmospheric Environment*, 30, 1209-1220, [https://doi.org/10.1016/1352-2310\(95\)00442-4](https://doi.org/10.1016/1352-2310(95)00442-4), 1996.
- Varon, D. J., Jacob, D. J., McKeever, J., Jervis, D., Durak, B. O. A., Xia, Y., and Huang, Y.: Quantifying methane point sources from fine-scale satellite observations of atmospheric methane plumes, *Atmos. Meas. Tech.*, 11, 5673-5686, <https://doi.org/10.5194/amt-11-5673-2018>, 2018.
- Witschas, B., Rahm, S., Dörnbrack, A., Wagner, J., and Rapp, M.: Airborne Wind Lidar Measurements of Vertical and Horizontal Winds for the Investigation of Orographically Induced Gravity Waves, *Journal of Atmospheric and Oceanic Technology*, 34, 1371-1386, <https://doi.org/10.1175/JTECH-D-17-0021.1>, 2017.