Thanks for giving us an opportunity to submit a revised version. We have improved our manuscript based on the constructive comments from the reviewers. Below you find one-to-one responses to all comments raised by the reviewers (our responses in blue). We are confident that we have addressed the concerns from the reviewers and hope that our manuscript will be accepted for publication.

RC2: 'Comment on acp-2021-1061',

Review of "Local to regional methane emissions from the Upper Silesia Coal Basin (USCB) quantified using UAV-based atmospheric Measurements"

General comments:

The authors present a manuscript following a measurement campaign in the Upper Silesian Coal Basin in Poland, where they used an AirCore attached to a UAV in order to measure CH_4 and CO_2 concentrations from several individual coal mine shafts. They make the case that a regional estimate derived from shaft-specific measurements will be superior to those that assign a single number to each mine, which broadly takes an average number across all shafts at any mine. They find good agreement between their methane measurements and high-resolution hourly inventory data in some of the shafts, whereas their flights were not able to reproduce well the coarser inventory numbers based on yearly estimates. They also claim that their CO_2 measurements have found that coal mines may be an overlooked source in the region that is not insignificant.

Overall, the paper is well-written and their ideas are clearly presented. The CH_4 analysis, in particular, is laid out in a straightforward manner that is easy to understand. That said, I feel that there are some important elements that are missing and some important changes that need to be made to this manuscript before it will be ready for publication. In particular, I have strong concerns about the CO_2 analysis, as I do not feel that there is enough data presented to support the conclusion that they have found a missing source of ~1% from the regional inventory, rather than it possibly being an artefact of upscaling. This, combined with my concern about how there is no independent value to compare the CO_2 measurements against, and some other smaller concerns I detail later, means my inclination right now would be to recommend leaving the CO_2 analysis out of this manuscript entirely.

AR: We thank the reviewer to point out the concern about the CO_2 analysis. First of all, we would like to clarify that the message was not that we have found a missing source of ~1% from the regional inventory, but rather we found a potential way of estimating CO_2 emissions from coal mine ventilation shafts and found it to be ~1% of regional CO_2

emissions. The upscaling to the regional CO_2 emissions is of course with assumptions, and we've tried to clarify this in the revised version.

Based on the strong correlation between observed CO_2 and CH_4, we derived the average CH_4/CO_2 slope of 4.6 $ppmCH_4/ppmCO_2$, which is consistent with the values found in Andersen et al., 2021. The derived slope and the quantified CH_4 emissions for individual shafts are used to calculate the estimated CO_2 emissions. On the shaft scale, the uncertainty of estimated CO_2 emissions is comparable to that of estimated CH_4 emissions. The upscaling of regional CO_2 emissions is tied to the upscaling of regional CH_4 emissions. We have rephrased the CO_2 part in the revised version to focus on the estimated regional CO_2 emissions and the uncertainty of the regional CO_2 emissions instead of a missing source of ~1% from the regional inventory.

We have removed the following sentence from the main text" Due to the omission of CO_2 emitted from underground coal mining in the E-PRTR inventory, we conclude that the CO_2 inventory is missing a source of roughly 1 %.".

I also may have found an issue with the stated method of how the hourly methane emission rate is calculated, for those numbers that they later compare their UAV measurements against and find relationships with, which I would like to see clarified.

AR: We have double checked the calculation of the hourly methane emission rate and found no issue with the calculation. Specific clarifications will be provided based on the individual comments below.

And I would also like to see further explanation/justification for the first of the three presented upscaling methods, which upscales the yearly inventory numbers using a relationship identified only with the hourly inventory.

AR: We have provided the requested explanation and we refer to our responses to the specific comment on Lines 407-412.

Further, I would like to see an expanded discussion of the possible areas of uncertainty including: i) the dangers of upscaling with such a small population of data (which they already do acknowledge briefly), ii) the uncertainties and potential misquantifications inherent in their plume calculation methods (the inverse Gaussian and especially the kriging), iii) the possibility of difficulties introduced by sampling at different times of day and under different atmospheric mixing conditions, and iv) at least some discussion of how the background was defined when calculating the leak rates, along with other ideas the authors may think of themselves.

AR: Thank you for this list of discussion items. We have tried to address each, while staying within a reasonable length of the discussion by making use of our earlier publications where possible. We have added the following paragraph to the revised version:

“A detailed description of the uncertainty analysis for both the IG and the MB methods has been presented in Andersen et al. (2021). Here we only give a brief description. The uncertainty of the IG method is calculated as the standard deviation of a series of optimized emission rates generated by a large number of optimization runs (N = 1000). The uncertainty of the MB method is mainly determined by the uncertainty and the variability of wind speed and wind direction measurements.”

We have added the following sentence to acknowledge the small population of data: “We acknowledge that potentially large biases may have been introduced to the upscaling as the number of quantified shafts (5) is small compared to the total number of shafts (59).”

The minimum concentration of the entire flights was used as background, which was subtracted from the measured concentrations before calculation of the emissions for both the MB and the IG approach. The minimum concentration is not the same as a typical choice of e.g., 10 percentile; however, the difference of the two values is relatively small compared to the large CH_4 enhancements, and thus causes negligible difference in the calculated CH_4 emissions.

We have added the following sentence: “The minimum concentration of the entire flights was used as background, which was subtracted from the measured concentrations before calculation of the emissions for both the MB and the IG approach. The minimum concentration is not the same as a typical choice of e.g., 10 percentile (Vinković et al., 2022); however, the difference of the two values is relatively small compared to the large CH_4 enhancements, and thus causes negligible difference in the calculated CH_4 emissions.”

I am curious, as well, as to whether the experimental set-up may mean that the AirCore samples are taken downstream of the rotors of the UAV, and whether that may introduce some dilution into their measurements (which may also help to explain why the measured values tend to be lower than the hourly inventory numbers).

AR: We appreciate this curiosity, and have been very careful ourselves in the design of this sampling system. The inlet of the AirCore system was positioned to the side of the carbon fiber box that is beneath the propellers. Therefore, the air sampled into the AirCore is effectively from above the propellers, within less than 0.5 m above the propellers (Lampert et al., 2020). As the UAV is most of time moving forward at a steady speed of 1-2 m/s, the collected air samples will not be disturbed . The change of the effective sampling altitude for all transects on the order of 0.5 m has no significant impact on the quantifications.

We've added the following paragraph to the methodology section 2.1 UAV-based Active AirCore system in the revised version:

"The inlet of the AirCore system was positioned to the side of the carbon fiber box that is beneath the propellers (Andersen et al., 2021). Therefore, the air sampled into the AirCore is effectively from above the propellers, within less than 0.5 m above the propellers (Lampert et al., 2020). As the UAV is most of time moving forward at a steady speed of 1-2 m/s, the collected air samples will not be disturbed."

There are additionally smaller things that should be quicker to fix, but would also be essential, including double-checking the unit scale-factors on each figure that shows CH_4 mixing ratios (which often seem too high by a factor of 10) and the units on the CH_4/CO_2 ratios. I believe that starting with these changes will make a substantial impact on the quality of the manuscript, and that by the time it is ready for publication, it will be a valuable manuscript to the broader community.

AR: We understand that these high numbers confused the reviewer, but the CH_4 mole fractions were indeed that high, on the order of 100 ppm. We have double checked and can confirm the units presented in the manuscript.

Specific comments:

Lines 47-49: Is there a citation for the numbers in either of these 2 sentences? It's a key statement towards the motivation of the study—even referenced in the abstract—so I think it's important to show where those numbers come from.

AR: Thank you, we have added a reference "Saunois et al., 2020" for the numbers.

Lines 109-110: I am wondering how the AirCore was exactly "attached" to the UAV. I see that the AirCore is coiled up, and there is a reference a couple of lines up to "carbon fibre box housing". Is the AirCore contained within that housing, or is that just the housing of the electronics for the UAV? It would help if there was a picture showing the set-up. Particularly I am wondering how it was ensured that the AirCore was measuring from air that was undisturbed by the rotors of the UAV. It looks like this UAV has 4 vertical rotors, and if the AirCore is taking air from underneath (or otherwise "behind" the rotors), then there may be a risk that the rotors are mixing the air (potentially pulling in more dilute air from the background) just before measurement, and therefore affecting the measured mixing ratios. If so, I would be interested in knowing how much effect this may have on the ultimate measurements. And along those same lines, I would wonder about what the effect on sampling rate is when the UAV is moving, considering the primary driver of intake is the ambient pressure. (Is the AirCore exposed in a way that it would sample more when the intake is pointed towards the direction of movement, because of the higher pressure, and vice versa? If so, how might that affect the results?)

AR: The AirCore is contained within the carbon fiber box housing beneath the propellers (Andersen et al., 2018). However, the inlet of the AirCore is positioned to the side of the box. The air samples are pulled into the AirCore through a pump at the outlet of the AirCore. The pumping flow rate is ~21.5 sccm. Besides this, the change of altitude affects the inlet pressure and thus the sampling. The two effects have been taken into account in the AirCore retrievals (Andersen et al., 2018). The possible ram pressure ($=0.5\rho v^2$) is on the order of 1 Pa with a flight speed of 1-2 m/s through static air, which is similar to the pressure change caused by a vertical displacement of 10 cm, and can be neglected during the AirCore retrieval.

We have added the following paragraph to the manuscript:

"The inlet of the AirCore system was positioned to the side of the carbon fiber box that is beneath the propellers (Andersen et al., 2021). Therefore, the air sampled into the AirCore is effectively from above the propellers, within less than 0.5 m above the propellers (Lampert et al., 2020). As the UAV is most of time moving forward at a steady speed of 1-2 m/s, the collected air samples will be fresh."

Figure 2: The units on these colorscales seem at least a factor of 10 too high. Were the authors really detecting plumes of 150 ppm of methane?

AR: The CH_4 mole fractions were indeed that high, on the order of 100 ppm of methane.

Line 210: This methane sensor gives output as a percentage concentration? Am I understanding that correctly?

AR: Yes, the methane sensor indeed gives output as a percentage concentration.

Line 213: I might be misunderstanding this sentence. It says "about 5% of the vented air to the atmosphere is from air inflow via the ventilation shaft closure". I understand that to mean that there is some quantity of vented air in this region, and that 5% of that total ventilated air comes from the shaft closure here. That does not sound like the same thing as saying 5% of the total gas flowing through this shaft gets vented. In order for it to contribute 5% of the total vented gas, we would need to know what the total vented amount is, then we take 5% of that number and use that to see how much of the gas flowing through the shaft would have to be venting. So, have I misunderstood the statement here? If not, then the "95% of the flow-rate" scaling factor would not work.

AR: Sorry that our statement was confusing and apparently misunderstood. The measured air flow contains not only the vented air from the underground coal mines, but also some additions from the ambient from the ventilation shaft closure that accounts for about 5%.

We have changed the sentence to

"According to the statements of ventilation engineers, the measured air flow includes about 5% ambient air from the ventilation shaft closure, ..."

Lines 230-235: This is a lot of words to describe the math, and I think I got a bit lost. Would it be possible to include the simple formulas for these 3 upscaling techniques?

AR: Thanks for the good suggestion. We have described the three upscaling methods using equations as below:

$Q_{M1} = Q_{E-PRTR-regional} \times k_1,$

$Q_{M2} = \bar{Q}_{UAV-shaft} \times n,$

$Q_{M3} = \left(\bar{Q}_{hourly-shaft} \times k_2 + b\right) \times n,$

where $Q_{E-PRTR-regional}$ is the annual E-PRTR emission rate, $\bar{Q}_{UAV-shaft}$ is the mean quantified shaft emission rate, $\bar{Q}_{hourly-shaft}$ is the mean hourly inventory emission rate, k_2 and b are the slope and the intercept of the linear fit of shaft-averaged emissions between our UAV quantified and high frequency (hourly) reported emissions, while k_1 is the slope of the linear fit that is forced through zero, and n is the number of active ventilation shafts in the region.

We have added the equations to section 2.6 in the revised version.

Line 282: I'm not sure that I'm convinced that there is a potential difference between weekend/holiday and weekdays, given the mass balance numbers. The inverse Gaussian numbers seem more like they could suggest that, but is there a reason to trust these more than the mass balance numbers? Feels like one shouldn't hint at a conclusion either way. (I assume that the phrasing "this may indicate..." is maybe an attempt to stay neutral, but it still reads to me like it's leaning towards the conclusion that there is a relationship.)

AR: As mentioned in the reply to reviewer #1, we've removed all texts regarding the weekend/holiday and weekdays.

Figure 6: Maybe this shouldn't be explained in the caption, exactly, but I'm not finding where in the text it explains why certain flights were deemed worthy of a mass-balance estimate but not of a Gaussian estimate?

AR: The reviewer raises a valid question here. The two methods have been applied to all flights that fulfilled the criteria. The missing quantifications from the IG method for some flights are entirely due to failures of the optimization. Please also refer to our responses to the comment by reviewer#1 "A critical issue is how you address those flights where the maximum concentration is at the edge of the curtain".

We've added the following sentences at the end of the first paragraph in section 3.2 Quantified CH_4 emissions.

"Note that both the IG and MB approaches have been applied to all flights that fulfilled the criteria. The missing quantifications from the IG method for some flights are entirely due to failures of the optimization."

And the following sentence to the caption of Figure 6.

"The number of successful quantifications on each day from the two methods is indicated in the parenthesis."

Line 297: Instead of assuming, is there anyone who could be contacted/referenced that would have more insight into why this period is missing from the inventory data?

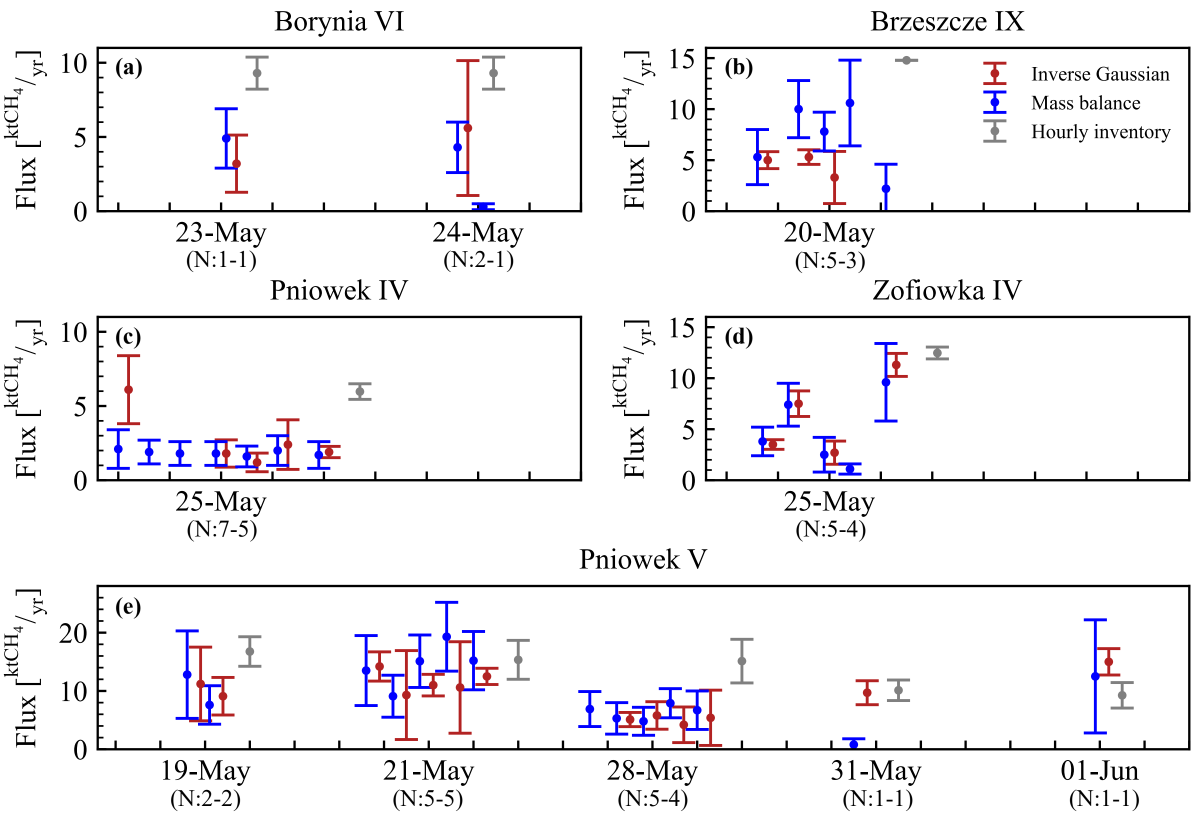
AR: Our collaborators in Poland, also co-authors of the manuscript, have tried to obtain more details, but unfortunately it is not possible to obtain more info.

Line 305: So the inventory seems to contradict the hypothesis that there's a difference between weekend/holiday and weekday emissions. To me, though, this seems consistent with the lack of conclusions we could have drawn from the data, anyway?

AR: As mentioned above, we've removed all texts regarding the weekend/holiday and weekdays.

Figure 7: Looking at Pniowek V, for example because it has the longest timeseries, the inventory would lead me to expect higher measured values on the 19th, 21st, and 28th compared to the 31st and June 1st, but that's not exactly what was seen in Figure 6, which shows low values recorded on all of the flights of the 28th and potentially high values on June 1st. Do we have an explanation for this discrepancy? (I actually think it might have been nice to combine Figures 6 and 7, so that we see the overlay of the measurements against the reported inventory directly.)

AR: Thank you, we have updated Figure 6 as suggested by the reviewer, see below. From the updated figure, we can see that the inventory estimates and the UAV quantified emission rates are broadly consistent within large uncertainties.



Line 313: Wouldn't we expect that the correlation between individual flights and yearly reported emissions would be very low, though? Because day-to-day variability would be so high, in comparison?

AR: Yes, the correlation between individual flights and yearly reported emissions would be expected to be low, and is in fact low according to our analysis.

Table 2: Could we convert this to a bar chart, maybe? (One could mark the max/min values separately from the error bars, and include the N numbers at the tops or bottoms of the value bars.)

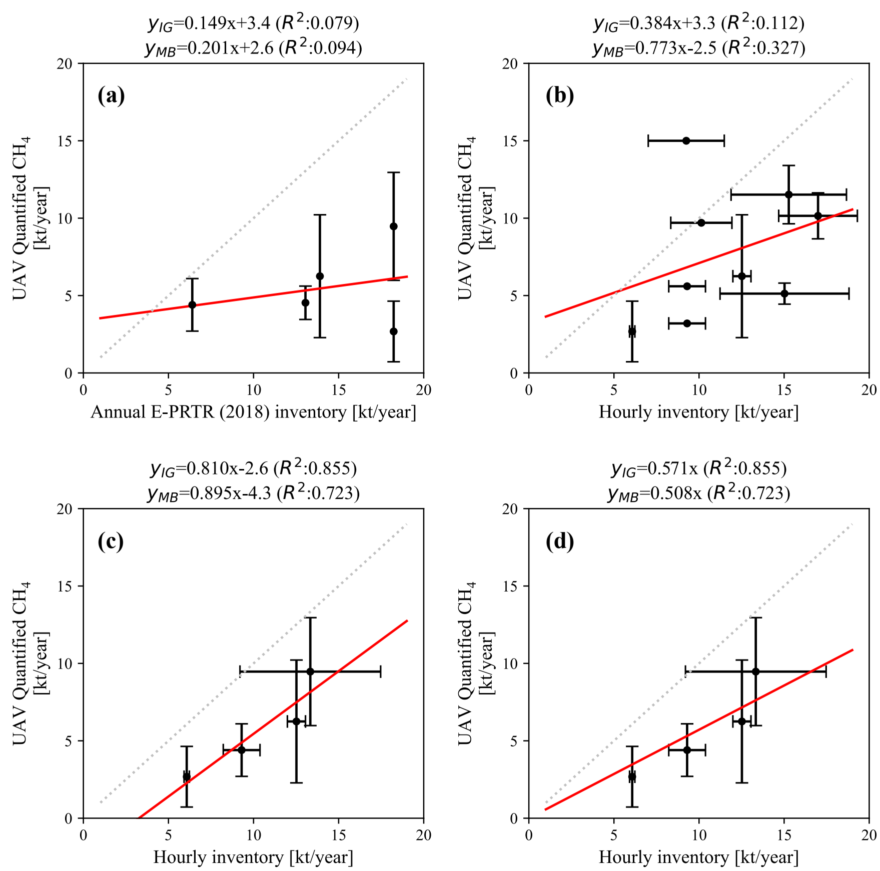
AR: Thank you for pointing out the duplication. The relevant information is already presented in figure 9. Therefore, we have moved table 2 to the supplemental info.

Figure 8: It's difficult to intuit where the 1:1 line would be with rectangular figures like this. Would it be possible to make these figures square with identical limits on the axes,

to really visualize the comparison? Maybe with a dashed 1:1 line, for reference? (I understand that this might necessitate dividing this up into 2 figures, in order to fit on the page.)

Figure 8: It also may be helpful to change the legends of each subplot to indicate that it's the best-fit line from the inverse Gaussian approach, specifically, as is noted in the caption

AR: Thank you for the helpful suggestion, we have indeed updated Figure 8, and added a dashed 1:1 line for comparison. See the figure below.



Line 331: What is the justification for forcing the linear fit through 0?

AR: This is to avoid of inferring an unrealistically large quantified estimate when the inventory estimate is approaching zero. We have provided two types of linear fits, with and without forcing the linear fit through zero. The linear fits for larger sources do not change as much as for near-zero sources.

Line 339: The hourly inventory is going to be used to scale up the UAV-measured concentrations?

AR: Indeed the reviewer is correct, the hourly inventory is used to scale up the UAV quantifications. Due to the relatively low correlation between the hourly inventory and the quantified emission from individual flight, we have used the correlation between the shaft-averaged hourly inventory and UAC-quantified emission rates to scale up.

Line 341: Of the linear fit from the multiple-days-averaged shaft-specific, inverse Gaussian case?

AR: We have used both the inverse Gaussian and the mass balance estimates for scaling up, and have changed the sentence to "We use the slopes and the intercepts found in Figure 8c to scale up our quantified emissions."

Figure 9: Is this all the same info from Table 2, it seems? If so, maybe we can just get rid of Table 2 and refer to this instead?

AR: Indeed it is from Table 2, which is moved to the supplement based on the reviewer's earlier recommendation.

Lines 358-9: Does this also imply that the sample size might not be enough to accurately quantify the other sites?

AR: A good question indeed, but we can't really provide a specific number for the sample size that is enough to accurately quantify the emissions. More flights will certainly provide better statistics. We've removed the sentence "This indicates that this statistical pool is sufficient to accurately quantify comparable emissions".

Line 362: It doesn't look to me like there is overlap at Pniowek IV in the mass balance approach…?

AR: True, we've removed "the mass balance approach" from the sentence.

Lines 374-5: I think here is where to mention the possible explanations for lower quantification in the air than what the hourly measurements within the shaft show, rather than lines 362-364, which was specifically talking about Pniowek IV

AR: Thanks for the suggestion. We have moved the possible explanations in lines 362-364 to lines 374-5 of the original version, and changed the sentence to "This could be due to a lack of statistics in the number of quantifications or the possible biases of the measured hourly inventory"

Line 382: One thing I don't think I understand is, if CO_2 has been measured as well as CH_4 from the AirCore, then why not just calculate the emission rate of CO_2 in the same way as was done with CH_4? Why introduce some linear dependence with methane and throw away the data that does not sufficiently have that linear dependence? Is the thinking that, if there are enhancements seen with CH_4, then it's presumed to originate from the shaft, but if there are enhancements in CO_2, they could also be from elsewhere nearby (are there other CO_2 sources nearby, like running engines?)? So this is done in order to ensure that one only looks at CO_2 that is believed to be from the shaft?

AR: We've used the linear correlation between enhanced CH_4 and CO_2 to calculate the CO_2 emissions instead of directly using the CO_2 data for two reasons: 1) the CO_2 signal is relatively small compared to its variabilities, which makes it difficult to find a robust background signal; 2) it is indeed as the reviewer mentioned that we aim to quantify the CO_2 emissions from the shaft only.

We have added the following sentence to the end of the paragraph in the revised version:

"We've used the linear correlation between enhanced CH_4 and CO_2 to calculate the CO_2 emissions instead of directly using the CO_2 data for two reasons: 1) the CO_2 signal is relatively small compared to its variabilities, which makes it difficult to find a robust background signal; 2) we aim to quantify the CO_2 emissions from the shaft only."

Line 383: The authors probably should specify which is the numerator and denominator in "slope", even if it seems obvious.

AR: Thank you, we added "(CH_4/CO_2)" at the end of the sentence "the slope is the slope of the linear fit between CO_2 and CH_4 (CH_4/CO_2)".

Lines 385-6: Would it be possible to include these scatter plots in the supplemental info, as well? I'm curious to see what they look like.

AR: Yes, we've added the scatter plots in the supplemental info, 5.2 Scatter plots of CH_4 and CO_2.

Line 387: I'm assuming the units are supposed to be ppb/ppm and not ppm/ppm? Additionally, this caused me to look at the figures in the supplemental info, where the flight tracks are provided, and it looks like the scaling factor on many of the colorbars is listed as 10^4, but it should be 10^3, since background methane should only be around 2ppm, not 20ppm.

AR: It is indeed ppmCH_4/ppmCO_2. The CH_4 mole fractions were indeed that high, on the order of 100 ppm of methane. However, the background was around 2 ppm, instead of 20 ppm, which can be seen from the updated color scale in Figs. S1-4 with updated colormap.

Line 392: Can the authors explain the NaNs again here? If there's not enough data to include an upper and lower bound, maybe it's better just to state that than to present it as a NaN value.

AR: The error bar is indicated as NaN when only one estimate is available. We have added it to the caption of figure 10.

Lines 407-412: I don't think I'm following the logic here. Figure 8 showed that there was no clear linear relationship between the measurements and the E-PRTR inventory, but that a relationship may instead be found when comparing against the hourly inventory. Then, here, the linear relationship that was found between the hourly inventory and the measurements is used to scale the E-PRTR inventory? What's the rationale for that?

AR: We acknowledge that this part is easily confusing for a reader, and we tried to explain it better in the revised text.

The E-PRTR inventory for each coal mine was provided, but the E-PRTR inventory for individual shafts were obtained by dividing the inventory for individual coal mines by the number of active shafts, which may have introduced large errors and may explain the very low correlation between the shaft-averaged E-PRTR inventory and UAV-quantified emissions. Here we scale the E-PRTR annual inventory for all shafts, assuming that the correlation between the shaft-averaged hourly inventory and UAV-quantified emissions are representative for the whole basin.

We have added the following sentences to section 3.5 Upscaling to regional estimates in the revised version:

"Here we assume that the correlation between the shaft-averaged hourly inventory and UAV-quantified emissions are representative for the whole basin and that the very low correlation between the shaft-averaged E-PRTR inventory and UAV-quantified emissions is mainly due to large errors introduced to the E-PRTR inventory for individual shafts by dividing the inventory for individual coal mines by the number of active shafts."

Lines 414-419: Might want to include an acknowledgement that the number of sampled shafts is small compared to the total number of shafts in the region (and among those sampled, those that have a large number of samples is even lower), so they may not be representative of the region as a whole.

AR: We fully agree, and have added the following sentence to the revised version:

"We acknowledge that potentially large biases may have been introduced to the upscaling as the number of quantified shafts (5) is small compared to the total number of shafts (59)."

Lines 421-2: My comment from the last paragraph should apply here, too. Though I think this is a much more sound approach than the first approach (which I would be tempted to toss out altogether without a clearer justification for why the hourly linear relationship would be directly applicable to the E-PRTR estimates).

AR: As explained above, we assume that the very low correlation between the shaft-averaged E-PRTR inventory and UAV-quantified emissions is mainly due to large errors introduced to the E-PRTR inventory for individual shafts by dividing the inventory for individual coal mines by the number of active shafts.

Line 441: When saying that they aren't statistically different when factoring in the uncertainties, should probably also acknowledge that the uncertainty bars are around 30%, which can be quite large.

AR: We fully agree and have added the following sentence to the revised version: "…, although the uncertainties are as large as 26-45%."

Lines 448-460: This illustrates the danger of upscaling to a region from just a few measurements. The authors note that coal mining activities are not a major source of CO_2 in the region, and that their measurements are also very low. The flight paths for the CO_2 enhancements are not included, so it's not apparent how clear or strong the CO_2 plumes really are compared to the background. Although Figure 10 shows that, though many of the quantifications do not have error bars, the ones that do are often quite large (e.g. Pniowek IV and Zofiowka IV). And since the E-PRTR inventory does not include coal mines in their inventory, there appears to be no way to independently check whether these values correspond to what would be expected or not.

AR: Based on our response above, we believe that the correlation found between observed CO_2 and CH_4 enhancements are strong and it is convincing to obtain the CO_2 emissions based on the estimate of CH_4 emissions. Also, because the CH_4 enhancements are very large, and the CO_2 enhancements are relatively small compared to its variabilities, the coal mine related CO_2 emissions can only be obtained through the linear relationship.

Lines 475-476: I do not think that one can conclude that the CO_2 inventory is missing a source of about 1%. Without having more information presented about the nature of the CO_2 plumes that were quantified, it seems within the realm of possibility that contemporaneous CO_2 data recorded with the CH_4 data displayed some stochastic variations (especially if the atmosphere is not well-mixed) that could mistakenly be quantified as small plumes with the inverse Gaussian or kriging techniques, especially if the corresponding background values are not well defined. Then, by scaling up those small numbers to the size of the region, they become an apparently large number (~1%). But this feels to me more like a potential artefact of the upscaling than a real missing piece of the inventory. Would we otherwise have any reason to expect large amounts of CO_2 to come out of coal mines? (If so, this is something that I guess should also be addressed in the introduction?) Overall, it is starting to feel like it may be best to leave out the CO_2 analysis altogether.

AR: Besides what we have responded above, there is indeed a reason to expect large amounts of CO_2 from coal mines. According to Swolkien, 2020, CO_2 emissions accompany CH_4 emissions during the extraction of coal.

We have added the following sentence to section 3. 5 Upscaling to regional estimates.

"According to Swolkien, 2020, there are collocated CO_2 emissions along with CH4 emissions during the extraction of coal."

And have added the scatter plots of CO_2 and CH_4 enhancements from the flight measurements to the supplemental info.

Line 496: I thought it was only this large for 25 of the 36 flights? And again I think these units are incorrect.

AR: Thank you for spotting this, as there was indeed a mistake here and it should be 25 out of 34 flights. We have corrected it in the revised version. However, the unit is correct.

Lines 509-511: I really disagree with this conclusion without some compelling evidence that it's not just an artefact of the upscaling.

AR: We think the disagreement partly stems from the phrasing, which we discuss also in response to reviewer #1. Our intention was to present a way of estimating CO_2 emissions from coal mine ventilation shafts and put them into context, which is why we mention it to be ~1% of regional CO_2 emissions by other methods.

Line 516: Maybe the authors should point out that their data indicated that at least 5—and probably more—good flights were needed for a decent quantification of a single shaft.

AR: Thank you, a good point to add. We have added the following sentences to the revised version:
"The uncertainty of the estimate of individual shaft can be reduced by increasing the number of the quantification flights, although it is challenging to determine the exact number of needed flights to achieve certain uncertainty. Analysis of a large number of controlled tracer release experiments may provide an opportunity to directly address this issue, as has been performed for UAV measurements as well as many other different measurement platforms (Feitz et al., 2018; Bell et al., 2020; Morales et al., 2022). "
Lines 513-520: All of this (good) assessment of uncertainties should have, I think, belonged in the discussion section. It's fine to repeat it here, but it felt like it was lacking above. Additionally, included in the discussion of uncertainties should be a discussion of the inherent uncertainties involved in the techniques applied (especially with kriging, which can be a very uncertain way to quantify a plume!).

AR: We have moved the sentence "The uncertainty in the emissions quantified by UAV-based AirCore measurements is linked to the stability of the wind, as discussed in Andersen et al. (2021). The 10-12 minute snapshots are not instantaneously sampled, and an unstable wind may cause the emission plume to meander across the plane. " to section 3.2 Quantified CH_4 emissions. And the discussion on the uncertainties of the quantification has been added in section 2.4 Emission determination.

Technical Corrections:

Figure 3: One of the labels is cut off—the one attached to the red marker.

Done

Line 241: The isotope numbers should be in units of permil, not percent. It's correct in the figure, but not in the text. (May need to be corrected throughout the manuscript.)

Done

Line 300: "emitted emission" seems redundant

We have changed "emission" to "CH_4"

Line 328: "on an hourly basis"

Done

Line 332-3: “not significant”

Done

Line 338: “Our evaluations indicate”

Done

Line 370: “all overlap with”

Done

Line 374: replace “more than one flights” with “multiple flights”

Done

Line 381: remove “emission” from “emitted CO_2 emission”

Done

Figure 10: The caption describes these plots as “histograms”. I do not believe that’s the case.

Changed to bar plots

Line 403: “As many as”

Done

Line 448: “linear curves” should be “linear fits”

Done

Line 490: “show very low…”? agreement? correlation?

Added “correlation”

Lines 526-28: This last sentence feels like a long fragment instead of a complete sentence, and should probably be reworked

We have changed the sentence to “The UAV-based active AirCore system can be a valuable tool to estimate CH_4 emissions on local to regional scales.”