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.

RC1: 'Comment on acp-2021-1061', Anonymous Referee #1, 24 Jan 2022

General Comments:

In this paper, Andersen et al. use an AirCore UAV system to quantify CH4 and CO2 emissions from coal vent shafts in Poland using two different methods. These quantified results are compared with directly measured (in stack) emissions, both hourly and aggregated by vent/day. Andersen et al. then use various techniques to upscale the quantified emissions in order to obtain regional emission estimates.

Overall, this work has important implications in the field of methane quantification from coal vents. However, there are non-negligible gaps in the science presented here. Most notably, the authors do not adequately explain the effect of and error introduced by flights when the maximum detected concentrations are on the edge of the "curtain", indicating that the peak plume concentrations may not have been sampled. Furthermore, the manuscript lacks an important figure which directly compares each flight-based quantification with direct (in stack) measurements. The manuscript can also be improved with a clearer, more coherent argument for the potential impact the manuscript will have on the state of this science. With these changes, this manuscript will be a valuable addition to the literature surrounding methane quantification, AirCore viability, and upscaling procedures. I was glad to get to read this manuscript and provide (hopefully) helpful feedback.

AR: We are pleased to see the encouraging comment. As the concerns are also raised in the specific comments, we respond to the concerns below.

Specific Comments:

I believe there is a lack of explanation in the methodology section regarding the quantification procedures.

AR: Yes, we have added a more detailed description of the quantification procedures. Please refer to our responses to the specific comments below.

For the inverse gaussian approach, what point(s) are plugged into the equation? Are multiple points used and compared/averaged?

AR: All the points collected by a single flight like the one shown in fig 2(a) are plugged into the equation (1). We did not average the flight data points but rather smoothed the simulated plumes according to the spatial resolution of the AirCore measurements. See also our response to the question below about the effect of AirCore smoothing on the quantification.

Is the maximum concentration used and assumed to be the center of the plume? How are the dispersion parameters determined (what method), and how do they affect results?

AR: No, we do not use or assume the maximum concentration to be the center of the plume. The dispersion parameters in the horizontal and the vertical direction (σ_y and σ_z), together with the emission rate (Q) and the coordinates of the center of the plume in the curtain (height H and distance D) are five unknown parameters that are optimized using the following method.

We have added the following paragraph in the revised version (the last paragraph in section 2.4):

"The AirCore flight data (Y) presented in fig 2(a) is compared with the plume simulations of the Gaussian dispersion model. A best fit for eq. 1 to the data can be found for these five parameters by minimizing the cost function $J\left(Q, \sigma_y, \sigma_z, H, D\right) = \left(C\left(Q, \sigma_y, \sigma_z, H, D\right) - Y\right)^2$ using a standard square error (SSE) approach. The five parameters include the dispersion parameters in the horizontal and the vertical direction (σ_y and σ_z), the emission rate (Q) and the coordinates of the center of the plume in the curtain (height H and distance D). A group of random starting points for the five parameters between their lower and upper boundaries are set for the optimizer each time, and the optimization is run 1000 times to ensure that not only a local minimum is found (Andersen et al., 2021). In this way, we obtain a series of optimized values for each of the four parameters as the final results, and the five unknown parameters are optimized simultaneously."

In fact, Andersen et al., 2021 already used the exact five parameters in the optimization scripts. However, the equation in Andersen et al., 2021 was written with four parameters without the distance parameter, D, which was a mistake that we only find out now.

Are concentration peaks dampened by the AirCore method due to mixing in the sampling tube before analysis, and how does this effect quantification?

AR: Indeed, the AirCore concentration peaks are dampened due to molecular and Taylor diffusions in the sampling tube, but mostly due to mixing of air samples in the cavity of the analyzer. Deconvolving the measured signal to obtain the unaffected concentration peaks is possible, as is done in Andersen et al., 2021. However, we have found that the

moving averages of the original data using an averaging kernel of 33–34 s can well match the convoluted signal. Therefore, the simulated data from the Gaussian model is smoothed with such an averaging kernel before comparing with the AirCore observations. This was thus performed for all flights during the data processing.

We have added the following paragraph to the revised version (the last paragraph in section 2.2):

"AirCore concentration peaks are dampened due to molecular and Taylor diffusions in the sampling tube, but mostly due to mixing of air samples in the cavity of the analyzer (Andersen et al., 2018). Deconvolving the measured signal to obtain the unaffected concentration peaks is possible, as is done in Andersen et al., 2021. However, we have found that the moving averages of the original data using an averaging kernel of 33–34 s can well match the convoluted signal. Therefore, the simulated data from the Gaussian model is smoothed with such an averaging kernel before comparing with the AirCore observations. This was thus used for all flights during the data processing."

A critical issue is how you address those flights where the maximum concentration is at the edge of the curtain. How are these flights interpreted? It is hinted at in section 3.2, but I'm confused as to how you are calculating either the IG or MB if the majority of the plume is outside the curtain. This may be clarified by some of the questions in the above paragraph. It would also be nice to see some type of error analysis for each quantification method; that is, how do things like wind variability, peak dampening, dispersion parameters, etc. introduce error and how is this error quantified.

AR: We have treated all flights that fulfill the sampling criteria presented in Andersen et al., 2021 in the same way. The criterion "the wind direction deviates from the vertical direction of the curtain is less than 15°" excludes flights where the maximum is outside the curtain and thus limits the number of flights where the maximum concentration is at the edge of the curtain. Furthermore, for some flights including those where the maximum concentration is at the edge of the curtain, the IG method fails because the optimization scheme won't find a solution. For the MB approach, biased results will be obtained because the missing information cannot be recovered, which means that our estimated emission rate will be underestimated.

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."

In the same vain, I think there is some issue with how error is represented in the aggregate data. For example, in the aggregation of quantified flux from Pniowek IV (Figure 5c) you claim an error of +-0.2kt/y due to the standard deviation of averaged points. However, in the individual day data for this vent (Figure 6c), the inherent error in each measurement is on the order of 3kt/y. A more robust error propagation analysis would make the aggregate numbers more defensible.

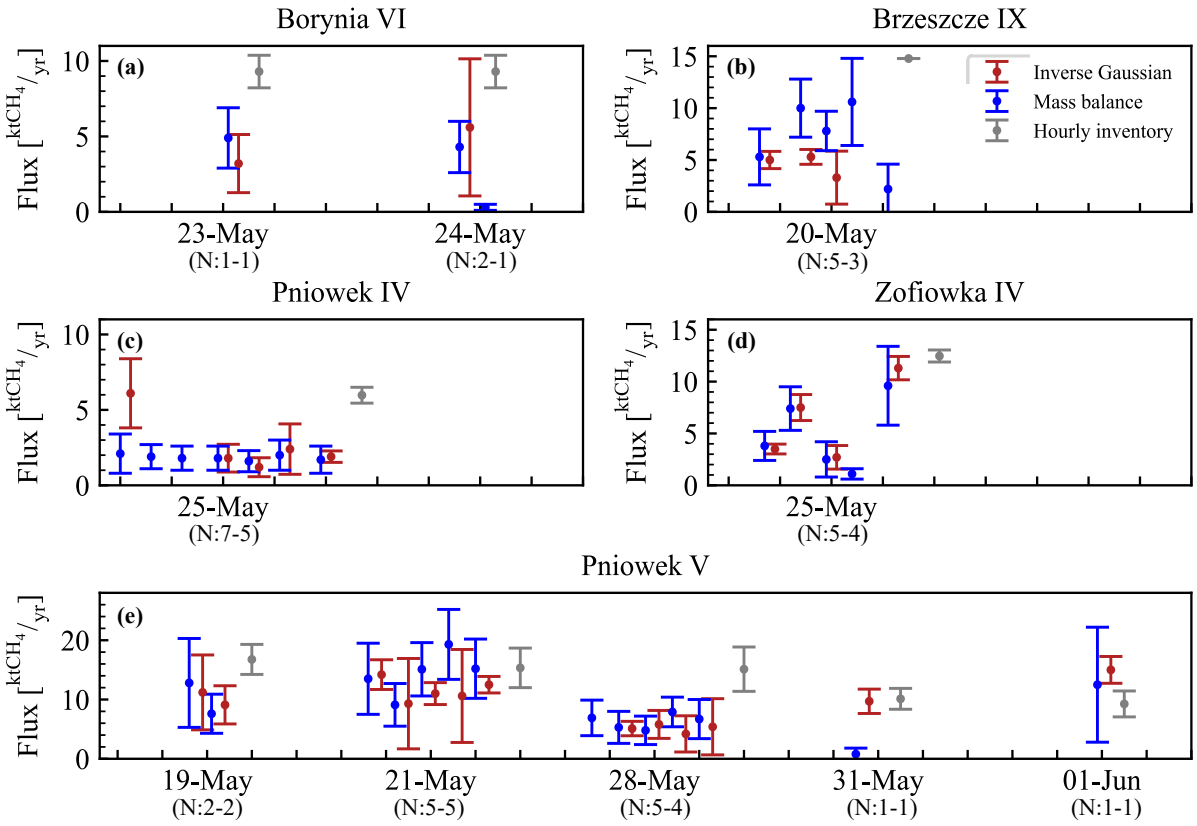
AR: In figure 5, we use the variability of multiple flights to indicate the uncertainty. Indeed, in figure 6, the error in each measurement is larger, because of systematic uncertainties that cannot be reduced statistically. We have discussed with a statistitian and got confirmed that the two types of errors cannot be simply merged as they represent different samples. We have therefore sticked to the use of the variabilitiy of multiple flights to indicate the uncertainty in figure 5.

A plot I'd really like to see is the hourly emissions compared with the flight quantified emissions (basically combining Figures 6 and 7). There is a bit of a roundabout comparison in the "hourly inventory" vs UAV quantified analysis (Figure 8b), but the critical representation is missing. The direct comparison is a key figure as it validates your UAV quantification approaches with real, empirical vent emissions data. As you state, emissions vary both intra and interday, so comparing UAV measurements at specific times with the directly quantified vent emissions instead of relying on aggregate data (like that presented in Figure 9) is an important distinction.

AR: We fully agree, and have added a direct comparison of the hourly emissions and the flight quantified emissions, and updated Figure. 6, see below.

We have updated the figure caption to "The quantified CH4 emission for each flight divided into different ventilation shaft and separated by individual flight days, with the hourly inventory. The emissions are also color differentiated by IG approach (red) or MB approach (blue). The number of quantifications on each day from the two methods is indicated in the parenthesis."

We have also updated the first sentence in section 3.2, "..., along with the hourly inventory presented the next section."



In section 2.1, you describe the del13CH4 data collection, but I am confused as how this is conducted. Are you capturing the outlet air of the Picarro upon measuring CH4/CO2 from the AirCore in a bag then analyzing? Some clarification would be helpful.

AR: Indeed, the air samples were collected at the outlet of the Picarro, downstream of the pump, and were stored in Tedlar bags for further analysis of isotopic signatures of $\delta^{13}C\text{-}CH_4$ and $\delta^{2}H\text{-}CH_4$ at a later time in the laboratory using a continuous flow isotope ratio mass spectrometer system.

We have modified and clarified the sample collection procedure in the main text in section 2.2.

Original:

"Directly after the CRDS analysis, the AirCore samples were collected in Tedlar bags for further analysis of isotopic signatures of $\delta^{13}C\text{-}CH_4$ and $\delta^{2}H\text{-}CH_4$. The isotopic composition was determined by analyzing the samples stored in the Tedlar bags using a continuous flow isotope ratio mass spectrometer system. More details about the analytical system and the calibration are provided in Brass and Röckmann, 2011; Röckmann et al, 2016; Menoud et al., 2021."

Revised:

"The AirCore samples were collected at the outlet of the Picarro, downstream of the pump, and were stored in Tedlar bags for further analysis of isotopic signatures of $\delta^{13}C$-CH_4 and $\delta^{2}H$-CH_4 at a later time in the laboratory using a continuous flow isotope ratio mass spectrometer system. More details about the analytical system and the calibration are provided in Brass and Röckmann, 2011; Röckmann et al, 2016; Menoud et al., 2021."

In my opinion, the argument of "weekends/holidays" does not add any value. If anything, it is confusing, as you postulate a reduction of emissions on these days then go on to show otherwise.

AR: We agree and have removed the "weekends/holidays" paragraphs in section 3.2 and 3.3.

Finally, I think there could be some added discussion about lessons learned and recommendations for future use of AirCore technology to quantify vent emissions. Specifically, expanding beyond *why* the hourly emissions data and individual flight quantifications may not align well and describing *how* the methods may be improved would be helpful. Similarly, some discussion of the recommendations for best practices to achieve a certain level of accuracy for quantifying vent/regional emissions using AirCore flights would be helpful; such as, how many flights are needed over how many days…etc.

AR: We thank the reviewer for the suggestion and have changed the conclusions section to "conclusions and outlook", and expanded the final paragraph. The paragraphs below have been added to the revised version.

"The use of UAV-based active AirCore measurements in combination with the inverse Gaussian approach and the mass balance approach has been demonstrated to be able to quantify the emissions from individual ventilation shafts, which can then be used to estimate regional emissions of both CH4 and CO2. However, the uncertainty of individual flight quantifications may be large, due to variable wind conditions under complexed turbulent schemes. Also, the in situ plume sampled by the AirCore does not necessarily follow the assumed Gaussian dispersion, as the averaging time is not sufficiently long, i.e., less than 30 minutes, which inevitably increases the uncertainty of the estimates by the inverse Gaussian method. To this end, optimization schemes that do not rely on the simple assumption of a Gaussian dispersion may be valuable (Shi et al., 2022). On the other hand, the complexed dispersion of the plume can be simulated by 3D large eddy simulation (LES), which can provide guidance to the design of the sampling strategy and

help develop a suitable method to estimate the emission rates based on the in situ plume sampling (Ražnjević et al., 2022).

The uncertainty of the estimates of individual shafts can be reduced by increasing the number of the quantification flights, although it is challenging to determine the exact number of flights needed to achieve a target 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).

Also, the uncertainty of the regional estimates can be reduced by increasing the number of quantified shafts. The limited number of quantified shafts makes our upscaling to the regional emission vulnerable. Nevertheless, the UAV system is flexible and versatile, and opens up opportunities to quickly obtain regional estimates in regions that are otherwise hard to access. The UAV-based active AirCore system, thus, has shown to be a valuable tool to estimate CH_4 emissions on local to regional scales."

Technical Corrections:

16: Insert (CH4) after methane.

Done

23: Delete "have"

Done

28: Delete "though"

Done

28-29: Rephrase "As an alternative…" sentence. Make sure verb tenses match and phrasing is clear.

Done

34: Is methane the second "most abundant" or just second most important in terms of climate forcing?

Changed to "most important". Methane is actually also the second most abundant anthropogenic greenhouse gas after CO2, while H2O is not anthropogenic.

49: Citation for coal being 12% of methane emissions?

Added a reference “Saunois et al., 2020”

52: Change “part of” to “some”

Done

54: Change “releases” to “is released”

Done

56: Insert comma between “mines” and “the”

Done

58: Citation for data loggers lacking accuracy and temporal resolution? It seems that your data shows otherwise… high resolution and temporally resolved fluxes from vents.

We have changed it to “lacking accuracy and continuity (Swolkień, 2020)”.

64: Sources for other studies using UAVs for methane monitoring?

We’ve had references using UAVs for methane (GHG) monitoring in the sentences following this one.

71: Perhaps add a line describing the Merlin mission and how CoMet ties in?

We have changed the sentence to “The CoMet aims at preparing the validation activities for the upcoming German-French Climate satellite mission MERLIN,” and added one more reference, Fix et al., 2018.

78: Change “strong ties to hard coal mining” to “ containing extensive hard coal mining” or similar.

Done

83: Period after PRTR

Done

83: Remove “the” after “quantify” and before “emitted”

Done

Paragraphs 70-100: Ensure consistent verb tense. Example: 71 – "goal of CoMet *is to* provide", 76 – "CoMet campaign *was to* quantify" etc.

Done

86/89: It goes from 59 flights to 34 quantifications – consider adding a line about filtering and what section you discuss this, otherwise it is confusing why these numbers don't match.

Yes, we have changed the sentence to "Here we present quantified emissions of 34 active AirCore flights that fulfill the flight selection criteria (Andersen et al., 2021)…"

89: The quantified emissions are of the shafts using the aircore, not quantified emissions of aircore flights.

We have changed the sentence to "Here we present quantified emissions of shafts using 34 active AirCore flights…"

95-100: Consider removing the "Section 2 presents …, section 3 contains…" and instead replacing with a strong statement about what your results convey and why they are important.

Yes, we have removed the "Section 2 presents …, section 3 contains… A conclusion is given in Sect. 4." We added "We show that a strong correlation (R^2 = 0.7 – 0.9) was found between the quantified and hourly inventory data-based shaft-averaged CH_4 emissions. Based on the correlation, we estimated regional CH_4 emissions by upscaling shaft-averaged CH_4 emissions. "

123-124: The names of the vent shafts have not yet been introduced and I did not know what these names meant. Consider revising to introduce the region and vent shafts before this section (maybe move section 2.3 to beginning of methodology).

We have moved section 2.3 to the beginning of the methodology section.

126: "First few": specify how many.

We have changed it to "first four".

126: What meteorological parameters were collected?

We changed the sentence to "...meteorological parameters (ambient temperature, pressure, relative humidity, wind speed, and wind direction) were measured using a radiosonde (Sparv Embedded AB, Sweden, model S1H2-R) identical to the one used in Andersen et al. (2021). "

129: Add "Meteorology for flights #5 through ..."

Done

Section 2.2: Add details about the height of the meteorological sensors.

We have added the detail "at about 1.5 m above ground."

135-137: "The CSAT3 has an operating temperature ... small changes in wind direction" is unnecessary.

We have removed the sentences.

144: Give some highlights about what the sampling criteria were to consider a "good flight"

We have added the details to the "the mean wind speed during the flight is larger than 2 m/s and that the flights are performed perpendicular to the wind direction (within 15°)."

144: The intro said 34 flights were used for quantification, this line says 36 fulfilled the criteria – why the discrepancy?

Changed to 34

146: Add "technique" between "this" and "effectively"

Done

153-154: Add the altitude range for the flight to go with duration and downwind distances.

Add "altitudes up to 100 m above ground" after "The flight duration varied between 8 and 12 minutes".

Figure 1: Is there any reason for the different colors for each vent shaft? If so label.

Just to distinguish the different mines.

179: How do you account for plume rise? In the gaussian equation, I believe his typically the “effective stack height” which accounts for advective or buoyancy rise effects of the plume.

The inverse Gaussian model is based on the assumption of steady state and can’t address plume rise. We have added the plume center height as a variable along with emission rate and dispersion parameters in horizontal and vertical direction to the optimizer that is described above. See also our responses to the general comment above.

Section 2.4: How is the local/regional background accounted for?

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 two values are close and do not add significant errors as the CH_4 enhancements are very large.

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.”

196: Add “estimate” after “annual emission”. Also, a source citing the E-PRTR inventory would be helpful.

We have added “estimate” after “annual emission”, and the reference to the E-PRTR inventory Gałkowski et al., 2021 was already there.

202: Add comma after “active shafts”

Done

210-211: How do you account for the fact that the operating range of the sensors is <100% RH, but the conditions are often over 100%?

We have removed the sentence “The conditions are often rough and the relative humidity is high, and the readings of relative humidity could exceed 100% when the filter is wet.”

215: Should “concentrations” be changed to “fluxes”?

No, here it should be “concentrations” instead of “fluxes”, and the emissions rate (flux) is given in the equation below.

243-244: The sentence “All the isotopic….” Does not make sense.

The sentence “All the isotopic signatures found from the UAV active AirCore flights” has been removed.

Figure 6: I’m confused by the color differences – did different flights use different approaches (MB or IG)? I thought each flight was analyzed in both ways? If not be more clear in section 2 about this. Label what the error bars represent. Consider making the x axis on (b), (c), and (d) so that there isn’t so much white space (restrict to sampling time period). Put in caption what the “N:7-5” means. Overall, I think there may be a better way to represent this data, consider reframing.

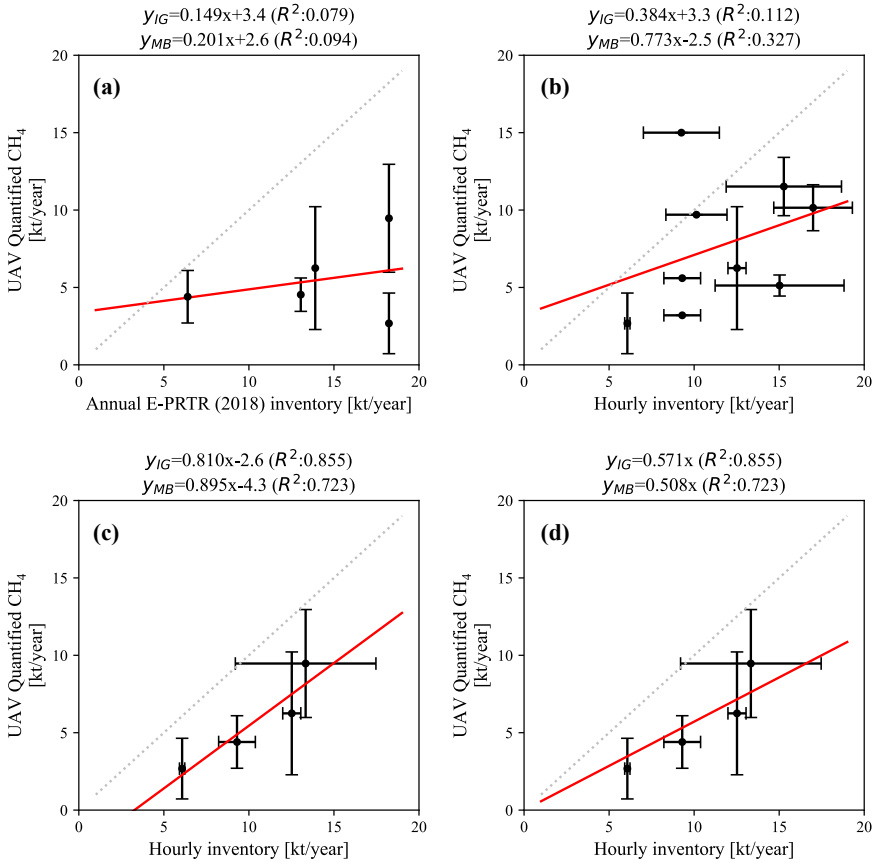
Both the IG and MB approaches have been applied to all flights that fulfilled the criteria. The missing quantifications from the IG method for some flights are entirely due to failures of the optimization. We have updated Figure 6.

298-299: “The Borynia VI inventory ‘may therefore not represent…’” I’d think it clearly does not, given the intra and inter day variability in your other data.

We agree. However, we do not have a better way of getting the inventory estimate for Borynia VI.

Figure 8: Many of the labels are obscured, overlapping, or otherwise can’t be read.

The figure has been updated. See below.



356: “Best statistics” – do you just mean the most flights? If so say that, if not clarify what “best statistics” means.

We have changed “best statistics” to “the most flights”.

363: Wording is confusing

We have 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.”

366: Again, is “lowest statistic” just fewest flights?

We have changed “lowest statistics” to “the fewest flights”.

370: “All over”? Confusing

We have changed it to “Thus, the measured distributions for Pniowek V, Pniowek IV, and Zofiowka IV overlap with the hourly inventory distributions”

421, 424, others: Replace “linear curve” with “line”

Done

421: Comma between “rate” and “calculated”

Done

450-456: Instead of “comparing” to estimates then talking about how the estimates don’t include coal, perhaps introduce this idea earlier. In reading, it is confusing why the numbers are so different until I realized that the EPRTR estimate really doesn’t represent coal emissions at all.

We have added the following two sentences before discussing the coal-related emissions. “According to Swolkien, 2020, there are collocated CO2 emissions along with CH4 emissions during the extraction of coal. However, CO2 emissions from coal mining activities are not included in the E-PRTR inventory.”

462: Add “method” after “upscaling”

Done

466/471: These lines contradict one another. “does not accurately represent emissions of the whole region” vs “a useful tool for regional emission estimates”. Best to clarify.

Here we meant to say that a simple approach by grouping the measured five shafts to obtain the average does not accurately represent the emissions of the whole region. However, using the slope and intercept of the strong linear correlation may provide a useful tool for regional emission estimates. This has been indicated in the paragraph “especially using the third approach of deriving the quantified emissions from hourly inventory data and scaling this to a regional emission rate”

500: Delete “have” between “we” and “used”

Done

S.I.: The color scale makes it so that the peak (and most critical part) of each plume is invisible (white).

Thank you for pointing out the issue. We have updated the colormap in Figure 2 and Figure S1-S4.