Reviewer 2

We would like to thank the reviewer for the helpful and detailed comments. In the following we carefully address the comments point by point. First we repeat the reviewer's comment (R#...) and then give our response (A).

R#1:The manuscript "Airborne remote sensing and in-situ measurements of atmospheric CO2 to quantify point source emissions" by Krings et al presents results from an airborne campaign, inferring point source fluxes of CO2 using both mass balance approaches as well as a Gaussian plume modeling of remotely sensed total column averages.
Even though the data presented here is indeed interesting, I tend to agree with reviewer 1 that it often reads much like a report and would need restructuring and more concise (and precise!) language.

A: We improved on the overall structure to meet the guidelines given by the two reviewers.

R#2: At times, the authors get caught up in details that are not entirely relevant to the study at hand, e.g. Figs 9-12 are too detailed or not necessary (9-10) or misplaced in the respective section

A: We removed Figures 6, 7, 9, 10, 17 and moved Figure 12 to the description of the target area, respectively. Figure 11, however, was left in the remote sensing section. Although it shows in-situ data, the plot specifically addresses the remote sensing analysis.

R#3: I would suggest putting a general description of the domain as well as the data right at the beginning (e.g. showing MAMAP footprints as well as in-situ ground projections on the map in Figure 1. It would greatly help setting the stage for the discussion.

A: Figure 1 has been updated accordingly.

R#4: Some more specific comments: Abstract last sentence: this is a sudden topic break and needs some rephrasing

A: We rephrased the abstract:
Reliable techniques to infer greenhouse gas emission rates from localised sources require accurate measurement and inversion approaches. In this study airborne remote sensing observations of CO_2 *by the MAMAP instrument and airborne in-situ measurements are used to infer emission estimates of carbon dioxide released from a cluster of coal fired power plants. The study area is complex due to sources being located in close proximity and overlapping associated carbon dioxide plumes. For the analysis of in-situ data, a mass balance approach is described and applied. Whereas for the remote sensing observations an inverse Gaussian plume model is used in addition to a mass balance technique. A comparison between methods shows that results for all methods agree within 10% or better for cases where in-situ measurements were made for the complete vertical plume extent. The computed emissions for individual power plants are in agreement with results derived from emission factors and energy production data for the time of the overflight.*

R#5: Page 3, line 30: wheras (there are many small things like this or "straight forward", which is one word. I won't go into more details, the copy-editor should catch those at a later stage but some sentences are too literal translations from german.

A: We carefully went through the manuscript to improve wording.

R#6: Section 4: I think this is poorly described and justified and I urge the authors to consider revisiting the differences between their approach and the paper Levi Golston mentioned.

A: This was done very extensively.

R#7: E.g.
I) you mention "Kriging" is not necessarily the best suitable approach. If you provide critic, you have to back it up either with an analysis or a citation. There is also no real explanation what kind of interpolation schemes you are using (apart from the boundary voxels).

A: As a main part of the revision, we did the whole calculations with our linear inter- and extrapolation method with only four rules, and with Kriging. Our method is now clearly described, and displayed in the new figures 4 to 8. There are two aspects to distinguish: (i) about the inter- and extrapolation. By using Kriging as another method we have shown that the difference is small. (ii) More important seems to average and interpolate fluxes, instead of averaging mass- and wind-fields before calculating the fluxes. Especially when the latter was done by Kriging, the deviation from the ensemble of other solutions is increasing, most likely due to the fact, that small artefacts in the individual fields are increasing the errors. Bottom line: The presented method is not better than Kriging, but, Kriging should not be regarded as the only option. This is also true after studying Gordon et al. (2015) in detail. The new text is much clearer in showing and discussing these details.
Since a complete revision was performed, and the separation of more individual sources was possible, the results as displayed in table 3 and Figure 18 were updated. The details are presented in a separate supplement.

R#8: II) You mention that you include turbulent fluxes. This is very interesting and I was excited but then I didn't see any further analysis. Did you compute the differences with or without turbulent components? What is the relative error in your case? Can you plot c-mean(c) vs v-mean(v) for some voxels to show the correlations as expected for turbulent fluxes?

A: Yes, this is done now. The different results can be found in a detailed Table as a supplement since it would be too much for Table 3 in the publication. The differences with and without turbulent fluxes were very small, and – surprisingly – negative, i.e. the turbulent contributions are rather from dilution (entrainment) than adding to the flux.

R#9: III) Wind speed seems to be a dominant error, do the others actually matter? You will need to provide realistic estimates regarding kriging and turbulent fluxes, otherwise the reader won't be able to judge the importance (even though this specific case study might not lend itself to extrapolation to a general case). Page 5, line 9: As above, please show what error you incur by doing mean(v)*mean(c) vs. mean(v*c)

A: This is now explicitly done by providing our standard fluxes (averages plus inter- and extrapolations of local mass x wind) plus 'flux 2', which was calculated after the averaging of the mass- and wind-field by our method, and by Kriging.

R#10: Figure 4: Please add color-bar and make this a realistic example based on real data.

A: The confusion about the old Figures 4 and 5 should now be eliminated. The new figures are much clearer and more consistent because it is clearly visible now which were the original measurements (Figs. 2 and 3), and how they were treated on the grids. This allowed to omit Fig. 4 which only showed the concept.

R#11: How many data points to average do you typically have per voxel?

A: A grid cell which was crossed once or twice is averaging the 5 Hz data (concentrations and wind), resulting in typically 10 to 20 data points when hres and zres were 100 m. However, the size of the cells was varied between 100 and 200 m for the sensitivity analysis, resulting in 10 to 40 points.

R#12: Page 9, line 8: Conditio sine qua non: Even though I have a "Grosses Latinum", I had to look it up again. Please rephrase in plain english, esp. as it is here used in a rather trivial way, not warranting the grandiose latin phrase ;-).

A: Was revised.

R#13: One might argue though that precision "could" be important if it is really bad while accuracy won't matter. This could be a factor when flying very cheap instruments on small unmanned aerial vehicles near the plume. So I would keep the discussions as general as possible.

A: We agree in principle. However, the absolute accuracy of the concentrations remains less important when subtracting the background based on the same measurements. This is also true for expensive and relatively stable instruments.

R#14: Page 9, line 17: Chimneys: It would be good to discuss how well you could measure the fluxes if the emissions were to happen at the surface. What would this imply for the in-situ based approach and potential flight-paths.

A: We discussed this in the referenced ESA report for CH_4, which was emitted from coal shafts close to the surface. Of course in such cases it is more important to have the lowest flight track as low as possible (50 m above ground with a special permission). However, in this campaign, where we flew to many sources for which we had no low-flying permission, this was restricted to 150 m. This was irrelevant for the CO_2 from high chimneys though. Another factor is the stability of the boundary layer. Since we have chosen daytimes for flying, when the convection reached at least the 150 mAGL, extrapolations with constant flux for the CH_4 (not in this paper), and diminishing or constant concentrations for the CO_2 (both options for checking the sensitivity) were applied. The constant flux for the CH_4 was applied because an expected enhancement of concentrations near the surface was compensated by the diminishing wind speed. For more reliable results for CH_4 we would prefer mobile measurements near the surface, which some of the authors did during a separate campaign in summer 2016. This aspect is now mentioned in the revised text.

R#15: Figure 5: This figure confuses me. I "assume" the dots are actual measurement locations.

A: See **R#10**.

R#16: Given what you wrote, there is a constant extrapolation to the surface. However, it doesn't look that way for the second little intrusion at x=0. Also, there are a couple of local maxima in between dots. You need to explain the interpolation scheme and this would a good place to compare against kriging or other interpolation schemes. Also, if the dots are measurements, please color-code them with the actual measurement values at that x-y position. This will help evaluate the interpolated fields better. A last point: Why is this continuous on x and y? Wouldn't it make sense to sketch out the actual grid boxes here as well?

A: Fully done. Figure 5 was confusing and misleading by several reasons: (i) Showing concentrations instead of the locally measured fluxes is potentially misleading; (ii) the old figure added interpolation & smoothing from the graphics package. The new style of cross sections in Figures 4 through 8 is much clearer.

R#17: Page 12, line 3: Please add citation for proxy method (this is not common knowledge).

A: We added the references to the revised manuscript:
See Frankenberg et al. (2005) and Schepers et al. (2012) for more information on the proxy method and Krings et al. (2011) for its application on MAMAP measurements.

R#18: Figure 6: I think the figure itself doesn't tell more than the text, could be skipped.

A: Figure 6 has been removed.

R#19: Same with Fig. 7

A: Figure 7 has been removed.

R#20: Page 14, line 17: So in essence, you don't really need the wind speeds in this case as the error is rather small?! Ideally, you won't always need both aircraft. If there is confidence in modeled winds, it would be a good sign for future remote-sensing only campaigns.

A: We agree with the reviewer. Not having to use additionally measured in-situ wind speed would indeed be a huge step forward. A wider and systematic analysis on the accuracy of the model wind would indeed be very interesting but can of course not be accomplished within the present study. We added a short discussion about this in the manuscript:

The analysis in this study shows an overestimated model wind speed of about 6% (or about 0.4 m/s) which is smaller than the uncertainty on wind speed. So in this case relying on the model alone may be sufficient. In a former study of similar setup (Krings et al., 2013) the error was about 10% or (0.7 m/s). A wider and systematic analysis on the accuracy of the model wind is needed to assess to what extent additional wind (profile) measurements are *dispensable. This will also become more relevant with regard to observations of localized sources by current and upcoming satellite missions with increased accuracy and spatial resolution. In these cases additional wind measurements will generally not be available.*

R#21: Figure 10: Weird x-spacing (value 493?). also better to use same x-scale for both subplots.

A: We removed the Figure completely as suggested in R#2.

R#22: Page 21, line 10 +/-: Wouldn't you ideally fit a Gaussian model with a vertical windspeed profile? This would rather directly model the total column AND the wind-profile.

A: Currently our model is set up to work with an average wind speed. The direct utilisation of the vertical wind profile u(z) for the inversion would be an interesting experiment. This would basically mean fitting the measurements to the sum of a vertically piecewise (or even continuously) changing Gaussian plume model (ignoring second order lateral and temporal wind speed variations in a first step). However, as Reviewer 1 pointed out that the discussion is "somewhat unnecessary", since in this case the vertical gradient in wind speed is not very strong we do not want to extend the discussion on this here and leave that to future work on the topic.

R#23: How high is your Gaussian profile extending to the vertical? That might be a plot to add (or is it just 2D in x and y?).

A: The Gaussian model to determine the average wind speed is 2D in along wind and vertical directions (x and z). With the same reason as above we do not want to add an additional plot for this but briefly discuss the vertical distribution as a function of downwind distance:
At the first remote sensing leg 700m downwind of power plant Niederaußem, the plume reaches about 1 km height, and at 2 km downwind distance the CO2 is already well mixed according to the plume model which represents a temporal average.

R#24: Page 23, line 19: "This is because they to a good extent cancel out..." "largely" cancel out?

A: Done.

R#25: Page 26, line 20: I think they gase don't need to be inert, just have lifetimes much longer than the time between emission and measurement. I would guess even NOx emissions could work with an "inert" assumptions on this very local scale.

A: Agreed. Has been clarified in the text of the revised version.

R#26: As a last general point: Please try to re-structure somewhat to bring out the key messages in a more concise way (and illustrate better how your in-situ inversions differ from others).

A: We agree and improved on the description as indicated in our answers above and to Reviewer 1.

R#27: At the end, provide a more generic overview of both flux estimates and

its pro/cons and path forward. This could extend to a discussion using high-resolution mapping like the cited AVIRIS-NG papers (which should be cited at page 26, line 9). Last but not least my sincere apologies for the late review.

A: Done. The discussion will be along the lines of what we also wrote as answer to Reviewer 1:

This case study illustrates the advantages and disadvantages of the used methods. The remote sensing approach offers the possibility to perform many flight legs in a short period of time. This is necessary to reduce the uncertainty as can also be seen from Fig. 11. The multiple transects allow for the application of the Gaussian plume model to a multi source setup which simultaneously retrieves the emission rates from several sources.

While for MAMAP the plume model usually utilises a priori information on the source location, an imaging instrument with sufficient spatial resolution and sensitivity (similar to, for example, AVIRIS-NG (Thompson et al., 2015; Frankenberg et al., 2016), though having a lower sensitivity compared to MAMAP) is able to determine the source location from the data directly and can acquire more data on shorter time scales potentially reducing uncertainties on derived emission estimates. Furthermore imaging instruments offer the possibility of mapping large areas in a survey for unknown sources.

However, there is generally the need for wind information which originates from models and/or in-situ measurements. The analysis in this study shows an overestimated model wind speed of about 6% (or about 0.4m/s) which is smaller than the uncertainty on wind speed. So in this case relying on the model alone may be sufficient. In a former study of similar setup (Krings et al., 2013) the error was about 10% or (0.7m/s). A wider and systematic analysis on the accuracy of the model wind is needed to assess to what extent additional wind (profile) measurements are dispensable. This will also become more relevant with regard to observations of localized sources by current and upcoming satellite missions with increased accuracy and spatial resolution. In these cases additional wind measurements will generally not be available.

The remote sensing instrument MAMAP measures solar backscattered electromagnetic radiation in the short-wave infrared. To simplify the radiative transfer calculations, cloud free atmospheres are selected to avoid the radiative transfer issues associated with solar electromagnetic radiation passing through clouds. The selection of the measurement day for this study was largely driven by this requirement. This generally involves a more convective and therefore thicker boundary layer making the gathering and analysis of the in-situ measurements more complex.

In contrast to the remote sensing measurements of the entire vertical column, in-situ measurements need to sample the plume with flight legs at different altitude levels. As a result of the time needed to complete a representative vertical cross section of measurements, only a limited number of repeated measurements are typically feasible. Interpolations within the cross sections and extrapolations to the surface and sometimes to the top of the plume have to be applied. This also applies for this study, where the boundary layer reached into restricted airspace. However, the in-situ method has the advantage of delivering vertically and horizontally resolved information in conjunction with co-located wind information, which can be readily used to infer a flux estimate. The high intrinsic sensitivity enables the detection of elevated trace gas levels also at great distances to the source. Errors on the inversion results from in-situ and remote sensing data are rather similar.

References:

Frankenberg, C., Meirink, J. F., van Weele, M., Platt, U., and Wagner, T.: Assessing Methane Emissions from Global Space-Borne Observations, Science, 308, 1010–1014, doi:10.1126/science.1106644, 2005.

Krings, T., Gerilowski, K., Buchwitz, M., Reuter, M., Tretner, A., Erzinger, J., Heinze, D., Pflüger, U., Burrows, J. P., and Bovensmann, H.: MAMAP – A new spectrometer system for column-averaged methane and carbon dioxide observations from aircraft: retrieval algorithm and first inversions for point source emission rates, Atmos. Meas. Tech., 4, 1735–1758, doi:10.5194/amt-4-1735-2011, 2011.

Schepers, D., Guerlet, S., and Butz, A.: Methane retrievals from Greenhouse Gases Observing Satellite (GOSAT) shortwave infrared measurements: Performance comparison of proxy and physics retrieval algorithms, J. Geophys. Res., 117, D10 307, doi:10.1029/2012JD017549, 2012.