In response to Anonymous Referee #1 comments from February the 5th, 2021.

The manuscript by Kostinek reports an interesting case study of top-down greenhouse gas emission quantification on the region scale using airborne observations, sophisticated atmospheric transport simulations and an inverse modelling framework. They estimate the CH4 emissions from coal mining shafts in Upper Silesia, one of the CH4 emission hot spots in Europe. The research topic is very relevant regarding greenhouse gas emission mitigation strategies under the Paris Agreement and as such deserves publication. The study design, the applied transport simulations and partly the inverse modelling approach are sound and mostly presented well in the manuscript. The authors took great care to assure adequate atmospheric transport simulations and made an important effort to characterise the connected uncertainties, adding some valuable new concepts to the field. There are two major concerns, 1) regarding the feasibility of assigning emissions to individual ventilation shafts given the flight observations taken at considerable distance from individual sources and 2) concerning the omission of CH4 sources other than coal mines. These issues are detailed below and an answer will probably require additional analysis and revisions of the manuscript. However, I would encourage the authors to address these additional points, so that their very valuable analysis can be published in ACP.

Dear Referee,
We want to thank you very much for this excellent review and the detailed, helpful comments on the analysis presented in the manuscript. We greatly appreciate your work involved with this review. The comments include very valuable concepts, that required additional analysis but ultimately improved the manuscript significantly. We hope to have answered your comments to your fullest expectation.

Major comments

Inverse modelling method: I have two major concerns with the emission estimation method.

1) Emission attribution to individual shafts The first concern is the attribution of emissions to individual facilities and shafts. Looking at figure 7 and 8 but also at the names of the shafts in figure 10, it is clear that many of the individual shafts cluster around individual mines at distances not much more than a kilometer. Furthermore, a lot of the locations actually line up with the main wind direction. In this situation it seems to be virtually impossible to estimate emissions from individual shafts from the presented observational data which was only taken at a single downwind curtain. The results presented in figure 10 are most likely a fine example of overfitting the observations and obtaining a 'noisy' a posteriori result. The problem is also apparent from figure 11, which seems to indicate that although there is some sensitivity to all emission shafts, sensitivity is much larger for certain shafts than for others. Given the observational data, the problem cannot be overcome in a general way, but at least the covariances in the inversion should be designed in a way that will limit overfitting and the overinterpretation of emission results. I would suggest that the authors modify the design of their a priori covariance. Currently, they only include diagonal elements. Hence, they assume uncorrelated uncertainties even for shafts from the same facility. I think it would be useful to explore by how much the results change if correlated a priori uncertainties for shafts from the same facility and/or shafts a shorter distances would be introduced (positive off-diagonal values in the a priori covariance matrix). Furthermore, the a posteriori covariance matrix should be explored in order to see if many of the emitters actually show negative covariances to one another. This would indicate that there remained a large uncertainty as to which shaft the emissions had to be assigned. In general, the manuscript should better highlight that the uncertainty on the shaft-level remained large and that the observational data is to limited for a more specific estimate. In this context, it would also be good to present the shaft level emissions on a map and compare with spatial inventories. This issue was also raised in the comment by J. Necki and he provides valuable discussion on a per shaft basis that should not be ignored in a revision of the manuscript.

Based on your advice we included "local information" as indicated in the short comment by Jaroslaw Necki into our a-prioris, by setting the a-prioris for the closed mining shafts to a very small number ($1\,\mathrm{kt\,yr^{-1}}$). In general we agree that our

Sect. 4.5 might have mislead the reader to think that we can accurately estimate individual shafts from two downwind walls. These estimates are certainly associated with significant uncertainty and correlations (as indicated by the inversion diagnostics), yet we are confident, that it is useful to include all available a-priori information into our state vector. With mission planning further optimized for the Bayesian inversion from airborne in situ data, as presented in this manuscript, these uncertainties can potentially be narrowed down in future campaigns. We rephrased Sect. 4.5 accordingly. In general, subregional emission estimates require at least two flights associated with different wind conditions. During single flights several emission sources are masked by sources closer to the point of measurement. In the present scenario, the large number of sources enables regional estimates for single flights, albeit subregional estimates will vary significantly and are not reliable for the reason mentioned above. For this reason we use two flights (morning and afternoon) that are close in time to reduce the impact of daily changes in emission rates yet allow for the determination of subregional emission estimates, albeit under the very challenging situation of only 10° mean wind direction delta between the morning and afternoon flights.

It is certainly true, that several shafts cluster around individual mines at distances not much more than a kilometer. However, that does not directly yield the amount of correlation of emissions from individual shafts. To our knowledge, methane emission strongly depends on the underground location where actual excavation is taking place. Furthermore individual shafts may also be used for fresh air supply. Therefore it is doubtful if the conceivable additional constraints (a-priori correlation) are appropriate or not. Yet we think it is a good idea to explore by how much the results change if correlated a-priori uncertainties for shafts from the same facility are introduced. The result for a $+.5$ correlation of the mines corresponding to clusters is shown in Fig. 1. As expected, the correlated a-priori covariance leads to a smoothing between individual shafts. The total emission

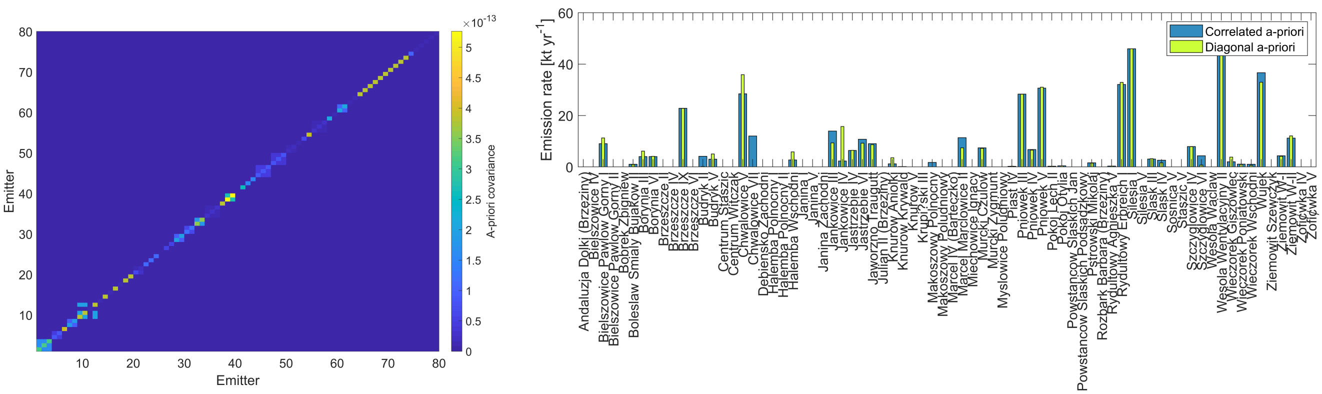

Figure 1. Left: +.5 correlated a-priori covariance. Individual clusters can be identified from the off-diagonal elements. Right: Impact of the correlated a-priori on emission estimates from individual sources.

rate changes by less than $< 0.5\,\%$, albeit there is some change in individual emission rates. In general the introduced additional a-priori information suggested by a-priori correlations remain to be verified, but we decided to follow this referee advice.

In contrast to the a-priori covariance matrix, correlations in the observational covariance matrix have a significant impact on individual estimates. The observational covariance had been estimated as a diagonal matrix in the first version of this manuscript neglecting correlations. In order to compensate for tempo-spatial-autocorrelation we simulated a puff release to check the impulse response of the simulation on a one-second release from individual sources. We sampled simulated observations from this puff release at the aircrafts location in space and time and computed the Autocorrelation function (ACF) from a single dispersed puff. The exponential decay of the correlogram suggested to augment the observational covariance matrix with a first order autoregressive model AR(1) structure with $\varphi = 0.7$. Fig. 2 shows a zoom on the first 400×400 elements of this 14936×14936 matrix to highlight the introduced off-diagonal elements. The modified observational covariance matrix led to a decrease in degrees of freedom for signal from $d_s = 48$ to $d_s = 32$, indicating that now more information stems from the a-priori. The right panel of Fig. 2) shows the a-posteriori correlation matrix as deduced from the covariance matrix. From the

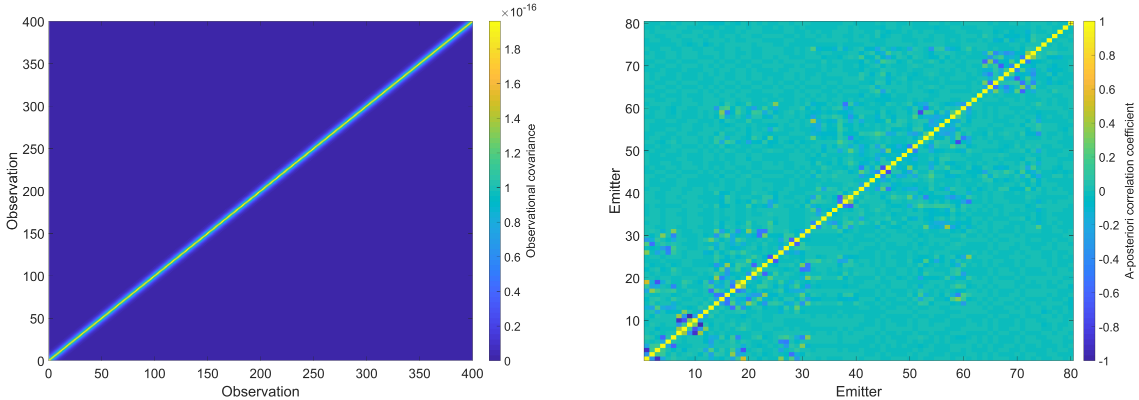

Figure 2. Left: Observational covariance augmented with a first order autoregressive model AR(1) structure with $\varphi = 0.7$ Right: A-posteriori correlation matrix deduced from the a-posteriori covariance matrix.

a-posteriori correlations matrix, there is some negative correlation as expected by the reviewer. In general, the uncertainties for individual shafts are enhanced as can be seen from the updated Fig. 3. This figure is to be compared to Fig. 10 in the original version of this manuscript.

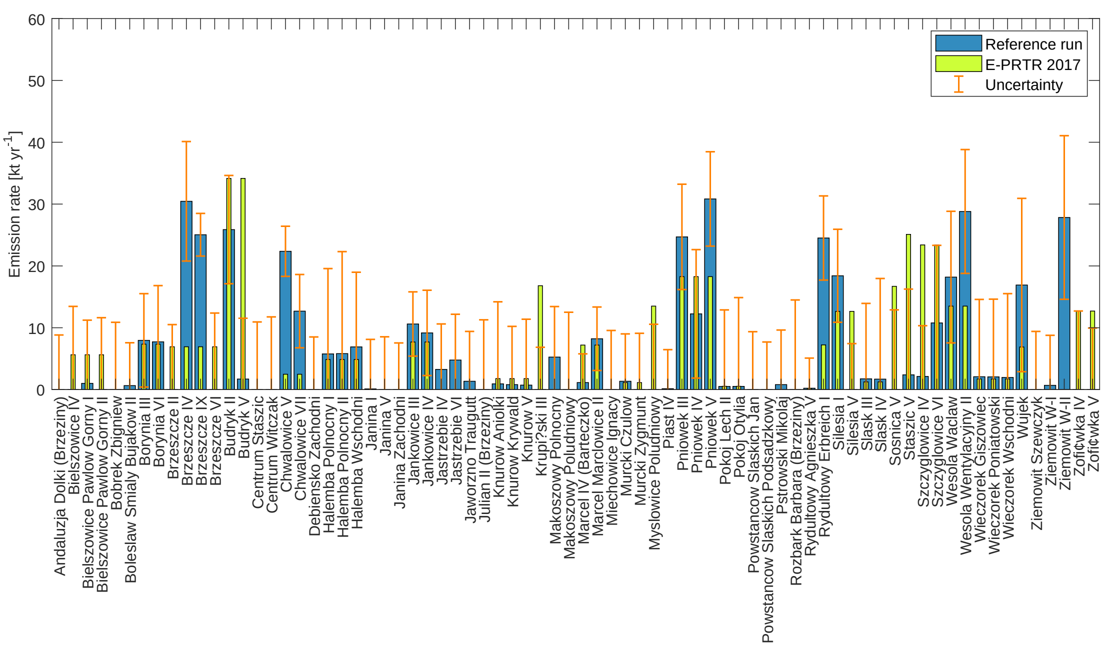

Figure 3. Emission estimates Φ_i (blue bars) in kt yr^{-1} for 74 individual mining shafts using the morning and afternoon flight of June 6th, 2018. Slim green bars are the reported yearly average values for each mining company (E-PRTR 2017) evenly distributed among the respective ventilation shafts. The orange error bars stem from the quadrature sum of the statistical uncertainties ϵ_i (computed from the parameter covariance matrix $\hat{\mathbf{S}}$) and the uncertainties $\sigma_{ensemble}$ derived from a variational ensemble with systematically perturbed parameters.

2) Neglect of other than coal mine emissions In section 2 it is discussed that EDGAR assigns about 14 % of total CH4 emissions in the area to non-coal mine emissions. The authors largely ignore these emissions in their analysis. Only in section 4.4 the possibility of emissions from the city of Krakow are discussed but ultimately these were also not included for the inverse modelling. Although 14 % may not seem a very large fraction and one could argue that any kind of mis-attribution is covered by the uncertainty estimates, it will still strongly depend on the distribution of these missing emissions relative to the sampling locations. Their impact on the concentration observations may well have been much larger if sources were closer to the flight track than the coal mine emissions. Since the authors also find that total emissions for the region are considerably smaller than what was reported in EDGAR, the fraction of non-coal may also be larger than in EDGAR. Seasonality of emissions may also play a role here. Since this is early summer, temperatures were above annual average, possibly leading to larger than annual average emissions from microbial sources like landfills, waste water treatment, manure management, etc. There seems to be a large discrepancy with EDGAR in terms of shafts and emission locations as well (red grid cell in the lower left corner of the encircled area close to Ostrava). Are these actually coal mine emissions in EDGAR? Looking at the area, one can also see a larger reservoir north-west of Bielsko Biala. Could natural emissions from this reservoir also play a role in the poorer agreement at the eastern end of the curtain flight? I strongly suggest that the authors reconsider their neglect of the noncoal mine emissions. They should obtain inventory data of these emissions, possibly not just from EDGAR (see comment below on inventories) but other more resolved inventories or from local information (see comment by J. Necki). With these emissions another FLEXPART forward run should be conducted and the resulting concentration either be removed from the observation vector before the inversion or an additional scaling factor for non-coal mine emissions should be introduced in the inversion.

The neglect of non-fuel-exploitation can arguably lead to the fuel-exploitation emission estimate to be biased towards higher values. Thus, additional partitioning into non-fuel and fuel-exploitation emissions does only further increase the projected discrepancy. As suggested, we re-analyzed our data, this time including non-fuel-exploitation sources. To this end, we added a figure to a revised version of this manuscript on methane emissions grouped by sector as provided by EDGAR to depict the spatial distribution of these sources and to showcase their small magnitude versus coal-mine emissions. The EDGAR v4.3.2 inventory includes information on sectorial partitioning of CH_4 emissions with non-fuel-exploitation making up for approximately $\sim$14 % of total annual CH_4 emissions in the USCB. From these $\sim$14 % approximately 90 % are attributed to the five sectors: Solid waste landfills, Energy for buildings, Waste water handling, Enteric fermentation and Oil refineries and transformation energy. Most emitters are weak in comparison to fuel-exploitation and uniformly distributed. Uniformly distributed emitters will be canceled out by the background subtraction and will hence not influence the emission estimate. Some stronger non-fuel-exploitation sources are however apparent. We added these source tiles (threshold $\geq 4\,\mathrm{kt\,yr^{-1}}$) to our FLEXPART-WRF simulation. This led to a partitioning of the emission estimates into non-fuel and fuel exploitation emission estimates for the USCB. The included non-fuel-exploitation sources are estimated with $27\,\mathrm{kt\,yr^{-1}}$ and $31\,\mathrm{kt\,yr^{-1}}$ for the morning and afternoon flights, respectively. This corresponds with the EDGAR a-priori ($33\,\mathrm{kt\,yr^{-1}}$) for non-fuel-exploitation to within 20 %. The derived fuel-exploitation emission estimates amount to $451\,\mathrm{kt\,yr^{-1}}$ and $423\,\mathrm{kt\,yr^{-1}}$ for the morning and afternoon flights, respectively. The small deviations of less than 5 % from the original run suggest a robust estimate for the USCB region. Nevertheless, including non-fuel-exploitation did make up for a substantial improvement, as this partitioning is only possible with sophisticated atmospheric simulations. This could lead to large errors in regions where non-fuel-exploitation is more relevant than in the USCB and/or for emission estimation approaches not involving atmospheric models. This lesson-learned will definitely be propagated to upcoming studies.

Minor comments

1. **p4: Regarding the use of the EDGAR inventory I would like to question if this is really the best available bottom-up inventory for the area. First of all, there is newer version of EDGAR available (v5.0 GHG), which explicitly lists CH4 from coal exploitation as a separate category and is available for a more recent year (2015) than EDGAR 4.3.2. Furthermore and as part of the EU project CHE, TNO has compiled higher resolution (6 km x 6 km) inventories for Europe. They may be better suited than EDGAR (see https://www.che-project.eu/sites/default/files/**

2019-01/CHED2-3-V1-0.pdf; data usually available on request). This is not only important for the final comparison of obtained emission estimates but also relates to the question if and how non-coal emission need to be treated in the inversion framework.

We have recently published a comparison between several available inventories for the USCB in Fiehn et al. (2020b) including inventories like E-PRTR, Scarpelli CH_4, CAMS-REG v3.1, EDGAR v5 and GESAPU. They all have their intrinsic advantages and disadvantages. Instead of re-iterating over the available inventories, we decided to showcase the large discrepancy between two well-known and well-established inventories to highlight the necessity of improving on bottom-up derived emission inventories via top-down GHG emission quantification.

2. **p5, l109: Here it is mentioned that an upwind concentration is subtracted from the downwind measurements. Later on a different method for background subtraction is described. What was really used?**
Some words were missing in this sentence, misleading towards the assumption, that upwind leg mixing ratios were subtracted from the downwind mixing ratios. We added the missing words: "[...] showing a fairly homogeneous CH4 inflow into the area of interest, thus allowing for subtracting an out-of-plume background (as described in Sect. 4.4) from the measured mole fractions downwind of the mines. [...]"

3. **p5, l113: Why is detrainment/entrainment important to this study? The FLEXPARTWRF simulations don't exclude detrainment/entrainment processes or PBL growth. Detrainment/entrainment would be more of an issue for a mass balance approach.**
This sentence is a historic residue, as the study was first based on a mass balance approach and only afterwards enhanced with high resolution particle dispersion simulations. We removed the obsolete sentence in a revised version of this manuscript.

4. **p5, l117: While an estimate of the morning PBL height is given, its height is not mentioned for the afternoon flight. Please add. Maybe also comment on the growth of the PBL height between the two flights and how this relates to the question of detrainment/entrainment.**
We added the missing information on the PBL height for the afternoon flight: "[...] During this flight, we observed an latitudinally inclined PBL with an approximate depth of $1.7\,\mathrm{km}$ a.M.S.L in the northern section and $1.3\,\mathrm{km}$ a.M.S.L towards the south. [...]"

5. **p7, L148f: Were the Doppler soundings the only observations that were nudged? What about other standard synoptic observations in the area?**
Yes, the Doppler soundings were the only observations fed into the observational four dimensional data assimilation. According to previous studies (see e.g. Cambaliza et al. (2014)) emission estimates obtained from airborne in situ data, are primarily affected by errors in wind speed. This is also apparent from our previous uncertainty analysis published in Fiehn et al. (2020a), where uncertainty in wind speed makes up for a large fraction of total uncertainty. For this reason, and because the used meteorological input data *NCEP GDAS/FNL 0.25 Degree Global Tropospheric Analyses and Forecast Grids* (GDAS/FNL, 2015) already assimilates global observational data, we refrained from nudging with conceivably the same observations again.

6. **p8, l158ff: Does this mean that 3Dvar and nudging were applied to the same observational data? That would not make sense in my view as the same information gets used twice. Rather use 3Dvar with smaller error covariance if the pull of the observations seemed too weak and such smaller uncertainties could be justified. Also, how were observational error covariances determined exactly?**
3DVar only assimilates observations temporally close to the 3-hourly available NCEP GFS analyses. As such 3DVar provides the initial conditions for the next 3 hours of WRF simulation. Meanwhile OBS-FDDA assimilates the data continuously during the WRF run as detailed in the manuscript. Smaller error covariances would not help. Instead, if analyses were available at smaller time intervals, the increased number of 3DVar runs could positively affect the results.

7. **p8, l167: Are these numbers the root mean square errors between model and observations for the 1-Hz sampling?**
These numbers correspond to the standard deviation of the residuals between simulated and 1-Hz sampled observational

data. We clarified this in a revised version: "[...] Simulated data, extracted at the aircraft positions in space and time, agree with 1 Hz observations of wind speed and direction to within an RMSE of $\pm\, 0.7\,\mathrm{ms^{-1}}$ $(1\,\sigma)$ and $\pm\, 5\,^\circ$ $(1\,\sigma)$, respectively. [...]"

8. **p8, l171f: How can this apparent offset in pressure be explained? Difficult to believe that the models (WRF nested in GFS) are off by that much, especially since the wind seems to match very well. Was there any comparison to other surface pressure data? Concerning the temperature offset: Does this vanish when you calculate potential temperatures? And same question as for pressure: were there any ground based measurements to compare to?**
We do not yet understand the reason for these offsets. The temperature bias does not vanish for potential temperatures. As described above, the Doppler soundings were the only observations used for data assimilation. We did not check with ground based measurements as these should be already assimilated in the meteorological input data *NCEP GDAS/FNL 0.25 Degree Global Tropospheric Analyses and Forecast Grids* (GDAS/FNL, 2015).

9. **p9, inversion method: I got confused by the description here. First, a non-regularized least square equation is presented for flux optimisation (eq. 1). Then regularization using a priori information and Bayes' theorem is advocated. To my understanding the resulting equations 4 and 5 only require a simple matrix inversion for solving for the a posterioir state. However, from line 211 onwards the application of a non-negative least square solver is presented. The latter is probably applied to equation 1, yielding a positive solution for x. However, if this was the case, I don't see why further analytical solutions to the cost function are presented in 4 and 5. I assume I am missing an important point here and would like the authors to clarify. If only a positively constrained solution for equation 1 is obtained I would think the results are even more overfitted as already mentioned above, since no additional a priori constraint on the individual sources would have been used. The description in the results section strongly suggests that this was the case. In equation, 4 I also think the last term Kx should also be Kx_a instead (see Tarantola eq. 3.37 or Jacob eq. 23).**
The introduction with a non-regularized least squares approach is intended to introduce and clarify the method. In the present case however, it is not necessary for the subsequent steps and has therefore been removed in a revised version of this manuscript. If the matrix $\mathbf{KS_aK^T} + \mathbf{S}_\epsilon$ is invertible, then a matrix inversion does the trick, albeit it finds a solution that also includes negative sources. These negative sources correspond to significant CH_4 sinks, and hence are treated as unphysical. The NNLS algorithm is used to discriminate sinks from the solution vector $\mathbf{x}$, yet it is not applied to Eq. 1 but to the MAP cost function Eq. 3 and therefore includes a-priori information on the emitters.
We revised the relevant parts of the manuscript to: "[...] Following a maximum a posteriori (MAP) approach, the scaling coefficients x_i can be found for each of the n modeled sources φ_i and for each of the m observed enhancements y_j making use of a-priori information $\mathbf{x_a}$ on the emissions of the individual shafts. Following Bayes' theorem the MAP solution is given by the minimum of the cost function (Tarantola (2004); Jacob (2007); Rodgers (2000))

$$J(\mathbf{x}) = (\mathbf{x} - \mathbf{x_a})^T \mathbf{S_a}^{-1} (\mathbf{x} - \mathbf{x_a}) + (\mathbf{y} - \mathbf{Kx})^T \mathbf{S}_\epsilon{}^{-1} (\mathbf{y} - \mathbf{Kx}) \tag{1}$$

with later defined a-priori and observational error covariance matrices $\mathbf{S_a}$ and $\mathbf{S}_\epsilon$, respectively. The MAP solution can be found by solving for $\nabla_x J(\mathbf{x}) = 0$ and is given by

$$\hat{\mathbf{x}} = \mathbf{x_a} + \mathbf{G}(\mathbf{y} - \mathbf{Kx}) \tag{2}$$

with the gain matrix

$$\mathbf{G} = \mathbf{S_aK^T}\left(\mathbf{KS_aK^T} + \mathbf{S}_\epsilon\right)^{-1} \tag{3}$$

By exploiting the averaging kernel $\mathbf{A} = \mathbf{GK}$ the number of degrees of freedom for signal d_s can be computed as

$$d_s = tr(\mathbf{A}) \tag{4}$$

This number describes the reduction in the normalized error on $\mathbf{x}$ introduced by the available observations and hence provides a measure for the improvement in knowledge of $\mathbf{x}$, relative to the a-priori, due to the observations.

The total emission estimate Φ in units $\mathrm{kg\,s^{-1}}$ follows from the scaled sum of the individual contributions φ_i

$$\Phi = \sum_{i=1}^{n} x_i \varphi_i \tag{5}$$

Here, the Non-Negative Least Squares (NNLS) algorithm (Lawson and Hanson, 1995) has been used to minimize the MAP cost function subject to the constraint $\mathbf{x} > 0$. This constraint is equivalent to the absence of negative sources. [...]"

10. **p16, l336: Here the total uncertainty of the emission estimate is presented as the sum of the 'systematic' uncertainty (which I assume results from the a posteriori covariance; eq. 8) and the spread obtained from the sensitivity simulations. Why are these uncertainty terms not added quadratically?**
As for any independent variables, the standard deviation of the sum is the quadrature sum of the individual standard deviations. We revised the respective text sections in the manuscript.

11. **Section 4.3: The way the covariance matrices for the a priori and the observation/model error are constructed most likely oversimplifies the true nature of the involved covariances and may lead to overfitted results. First, and already mentioned in the main comment above, the a priori covariance should acknowledge the fact that the a priori emission uncertainties will be correlated. This is true for shafts belonging to the same mining complex but may also be true for spatial distances between shafts. As mentioned above, I would suggest introducing off-diagonal elements in the covariance matrix to honor this fact. This would certainly lead to a smoothing out of the emissions across different shafts but is a more realistic approach. Furthermore, the observation/model covariance does not include off-diagonal elements either. However, the 1-Hz observations are certainly not independent from each other since they contain tempo-spatial autocorrelation. The latter will also be present in 1-Hz model residuals. I would suggest to explore this auto-correlation in the residuals and add a temporal correlation length to the observation/model matrix accordingly. Adding these off-diagonal elements will probably reduce the impact of the observations on the a posteriori results, reflecting that they are not really independent from each other. Another way to get rid of the autocorrelation would be temporal averaging of the observations before using them in the inversion. This has its merits as well as it would also bring the spatial resolution of the observations closer to those of the transport model.**
This comment is basically a detailed version of major comment #1. We replied above to major comment #1.

12. **p11, l231: How was the transport model uncertainty estimated concretely? As the standard deviation of simulated concentrations from the 8 ensemble members?**
The transport model uncertainty was estimated from 8 sensitivity runs, where the respective variables were perturbed in one or the other direction globally, but not perturbed by random noise with the given sigma width. Changes in the manuscript are described below at Comment #15.

13. **p12, l248: Why not use the measurements from the upstream flight segment as background? The comparison with the model output seems to indicate that the overall plume was wider than the flight segments.**
Due to the limited available flight time, the upwind leg is flown at a single altitude only. Subtracting the observations from this flight leg from the downwind wall pattern would require a Lagrangian simulation of the mixing ratioes during the upwind leg projected onto the downwind wall location. While possible, it would introduce significant additional uncertainty due to the single flight leg. For this reason we refrained from subtracting a Lagrangian propagated background and decided to use the out-of-plume background on both sides of the downwind wall.

14. **p14, l296: This sentence largely repeats the result from the previous sentence (EDGAR being much larger than the current estimate).**
The sentence is obsolete and has been removed.

15. **section 4.6: After reading the first few sentences, it was not clear to me how an uncertainty quantified by sigma was adopted in the transport model. I guess figure 11 makes it clear that 8 sensitivity runs were done where the respective variables were perturbed in one or the other direction globally, but not perturbed by random noise with the given sigma width. This should be made a bit clearer from the beginning.**
We revised the introductary sentences of this section to make the derivation of the systematic transport model uncertainty a bit clearer: "[...] The influence of several variables on the total flux estimate Φ has been computed from 8 sensitivity runs with symmetrically perturbed parameters. The systematic transport model uncertainty is subsequently estimated as the standard deviation of this ensemble. [...]"

16. **p15, l325: If I understand correctly, original horizontal wind speeds as output from WRF were increased/decreased by 0.9 m/s. In doing so, the local mass balance of the wind field may well be destroyed as vertical wind speeds were not adjusted (correct?). This may lead to errors in the transport description of the LPDM. Have you given this any thoughts? Probably the impact was not to large and since this is only presented as a sensitivity case it is of less importance, but it may have lead to larger discrepancies from the reference run than anticipated. A similar question for the PBL height. Is the latter taken from WRF or is the diagnostic calculation taken from FLEXPART? When increasing the PBL height just in FLEXPART vertical mixing in FLEXPART may then bring model particles to altitudes that in WRF are not part of the PBL and as such may have a distinctly different flow direction as flow in the PBL. As a consequence the differences to reference run may be larger than in a case where WRF PBL heights were larger/smaller. Hence, your change in the PBL height may give a slightly more pessimistic (larger) uncertainty.**
We are aware of, that by not adjusting the vertical wind fields, the local mass balance of the wind field may be jeopardized leading to larger residuals and hence larger uncertainties. However, as this is only a sensitivity analysis it is of minor importance here. In contrast, the PBL height has implications for all runs and hence also for the emission estimate. To retain a local mass balance of the wind field we do not use the diagnostic PBL height calculations from FLEXPARTWRF but feed the LPDM with the PBL height as simulated by WRF.
We added this info in a revised version of the manuscript: "[...] FLEXPART-WRF version 3.3.2 (Brioude et al., 2013) was used to model the exhaust plumes of known emitters forward in time using the meteorological data (including PBLH) obtained from the WRF simulations described above (see Sect. 4.1) as a driver. [...]"

17. **p16, l23: Here it is mentioned that the statistical uncertainty was estimated from eq. 7 and 8 and it is referred to elements e_i, which are the diagonal elements of the a posteriori covariance matrix. What about the off-diagonal elements of this matrix? Were the taken into account for the total uncertainty estimate?**
Off-diagonal elements are included for the regional emission estimate uncertainty. Including off-diagonal elements of the a-posteriori covariance matrix for single entries of the state vector would require a change into a basis where all off-diagonals vanish, if such a basis exists at all.

18. **Figure 12 and use of Jacobian: If I understand correctly, what is shown in figure 12 is the matrix K containing the elements dy_j/dx_i. However, the term Jacobian is also used in the manuscript for grad(J(x)). But J is not a vector-valued function and as such grad(J) is not a Jacobian. Please clarify.**
The Jacobian we are referring to throughout the text is the matrix $\mathbf{K}$. We removed the reference to grad(J) as the Jacobian in a revised version of the manuscript.

Technical comments

19. **p2, l45. Karion et al. is missing a publication date.**
The missing date has been included in a revised version of this manuscript.

20. **Figure 1: The line indicating the afternoon flight is more orange than red (as described in caption). It looks like the map is showing total EDGAR emissions. How does the distribution of non-coal mine emissions look like?**
We made the line in Fig. 1 look more reddish. The distribution of non-coal mine emissions has been included in a revised version of the manuscript.

21. **Figure 3: Please label the WRF domain in the figure according to their definition in the text.**
WRF domains have been labeled according to their definition in the text in a revised version of this manuscript.

22. **Figure 4+5: Please use the same colors for the different WRF runs. The legend is fairly small in both cases and needs to be enlarged, possibly put to the right of the sub-panels.**
Figure 4+5 have been adjusted accordingly in a revised version of this manuscript.