Author response to referee comments

We would like to thank both referees for their constructive comments and insight. They have provided us with plenty of food for thought, both in terms of the tone of the paper and its place within the wider context of the field. Below we respond to each referee individually, however as both referees raised similar points in many places we have combined our replies into a single document.

Referee 1:

Pg 1 ln 19-20. Consider rephrasing – many in the atmospheric greenhouse gas community are recognizing that the value of atmospheric measurements in the emissions reporting context is in working with existing inventories to evaluate and improve emissions reporting. Presenting these measurements as "independent verification" pitted against inventory methods is problematic.

A good point – amended to better emphasise the potential synergy between top-down emission estimates and bottom-up inventories

Pg 2 ln 3-10. Please add some discussion here about sources/sinks that are not included in the NAEI. For CO2 this is mainly biogenic fluxes, which are noted later to be critically important. Please also include a discussion of what sources of CO and CH4 are included in the inventory, and which are not. For example, I suspect that oxidation of biogenic VOCs is not included in the CO inventory. These may be negligible in March, but should still be mentioned. As for the CH4 inventory, does it include all sources, or only anthropogenic sources, and how significant might non-anthropogenic sources be?

A brief paragraph on the natural sources of these gases has been added. For CH_4 and CO the impact of natural sources on the results of this study is likely to be small, but as discussed later in the paper this is not the case for CO_2. Rather than repeat it here, we refer the reader to the relevant section for this discussion.

Pg 2 lines 21-22. While it is true that comparison of top-down estimates with bottom up inventories is one important way to use the atmospheric observations, it is certainly not true that the only use of these measurements is to evaluate inventories! Please rephrase.

We certainly didn't intend to imply that the measurements themselves were only useful for this purpose, only that the mass balance flux calculated through the downwind sampling plane is not particularly informative in isolation. This value, which physically represents the mol s^{-1} passing through some arbitrary vertical plane above some defined background mole fraction, needs to be related to the emissions for a given city/region/source for it to have any meaning. The lack of clarity in this section was also highlighted by referee 2 – it has been edited to address both comments.

Pg 4 ln 25. "an altitude-latitude plane."

Corrected

Pg 5 ln 6-20. Please add a sentence that explains in plain English the principle of what the equations do, rather than requiring the reader to wade through the equations to figure out the principle (although the detail of the equations is necessary too).

We've reordered this section to try and make it flow better.

Pg 5 section 3. Please add some detail about the NAEI. It is spatially explicit, but does it have temporal variability? If so, what kind of temporal variability and how reliable might that be? Diurnal cycles? Seasonal cycles? Weekday/weekend? Are there any existing estimates of the quality of the

inventory (and perhaps the quality is different for the different gases)? This becomes important in trying to understand the differences between the inventory and the observations.

The NAEI only contains annual averages – we've added a sentence to explicitly state this and alluded to the fact that this is likely to be a source of model-measurement disagreement for a single-flight case study.

Pg 6 ln 24-25. Again, please add a sentence that explains the principle in plain English rather than forcing the reader to work it out from the equations. Eg "The mole fraction enhancement is calculated by subtracting the background value".

Added

Pg 6 ln 24-32. The choice of background is known to be a key uncertainty in this type of measurement (eg. Cambaliza et al 2013; Heimburger et al, 2017). Unfortunately the research community has not yet come to any conclusion as to how to resolve this. The simple method of taking an average of the values measured on the downwind edges of the plume (as is done here) is far from perfect, even if it might be the best available option given the measurements that have been done. Heimberger et al (2017) showed that there can be significant differences in the values on the two edges, and that in that case, a simple improvement would be to linearly interpolate between the two edges to evaluate background. It is also entirely possible that the background is not uniform and that there are plumes from upwind sources that are not detected because they are inside the urban plume. From Figure 4, it's apparent that there are a lot of methane emissions upwind of the city that could cause this. Further, there's an implicit assumption that there are no emissions occurring in the footprint of the edge measurements. This is clearly a bad assumption for this dataset, and so the edge measurements will be biased high (or perhaps low in the case of CO2 if there is significant drawdown in the edges), resulting in an underestimate of the urban emission rate (or perhaps overestimate in the case of CO2). A forthcoming paper (in last phases of review) will discuss this further, but unfortunately is unlikely to be published in time to be referenced in this paper. My suggestion is to: (1) Add a figure that shows clearly the background values, how they were chosen, and whether there is any difference between the two edges. (2) A plot of the upwind measurements could also be included to show whether there is any particular concern with plumes coming in from upwind for this dataset. (3) Add figures that show the NAEI CO2 and CO emissions, similar to that shown in Figure 4 for CH4, to give a sense of upwind and edge emissions and how important they might be. (4) If there are no particular concerns with the points above, then stick with the current choice of background. (5) Add some discussion about the uncertainty associated with choice of background and how it might influence the results.

We absolutely agree with the referee that this is the key issue with the mass balance method – this is what motivated us to develop the flux-dispersion method which we think is less susceptible to the biases it can cause. The fundamental problem with applying the mass balance method here is that there are many emission sources outside London that contribute to mole fraction enhancements in the downwind plane. Hypothetically, to deal with this issue in the context of the mass balance method, one either needs to account for the influence of these emissions in the background mole fraction (such that all downwind enhancements are a result of Greater London emissions) or include them in the aggregated inventory emission total against which the top-down flux is compared. Our original draft focussed on the difficulties associated with the latter approach – we have redrafted Sects. 3.2 and 3.3 to include discussion on the former approach too.

One thing that is worth clarifying regarding the comment made here is that Pg 6 L24-32 describes the approach taken in the flux-dispersion method. However the issues raised here by the referee are more relevant to the mass balance method discussed in Sect. 3.2. The flux-dispersion method does take into account emissions in the footprint of the edge measurements, and it is made explicit in this section that the results obtained using this approach pertain to the emissions from the areas sampled in Fig. 4b relative to the emissions sampled in Fig. 4a.

We have repeated the flux-dispersion analysis with an interpolated (rather than averaged) background value used across each transect, and each ratio changed by less than 0.01. We have made reference to this in Sect. 3.1.1. Similarly we repeated the mass balance method with interpolated background values and found this slightly increased the derived fluxes, but in all cases the difference was less than 7%. We have included these values in Sect. 3.2.2.

On the specific points made above:

1) We have added a table containing the background values used in the flux-dispersion method as we thought this shows them more clearly than a figure.

2) Unfortunately the upwind sampling was nearly all either out of the boundary layer or dipping in and out of the top of it, so this data is not really useful for defining a background here.

3) These plots have been added to Fig. 4.

4) As the flux-dispersion method accounts for the emissions in the background footprint, the key to defining the background is to set a criterion that ensures emission from the area of interest (in this case Greater London) represent the most significant difference between the in-plume and background footprints. Comparing Fig. 4a and Fig. 4b we feel that the background choice here achieves this. This is discussed further in our response to Referee 2 below.

5) We have clarified the discussion in Sect. 3.1.1 of how the choice of background influences the flux-dispersion results, emphasising the impact of this choice on the selectivity of the results towards London. For the mass balance method we have added discussion on the choice of background to Sects. 3.2 and 3.3.

Pg 7 ln 24-30. Looking at figure 5, there's a clear spatial mismatch in the plume location between the obs and simulation. What might be the explanation for this? Given this mismatch, is it reasonable to average over the whole thing and then compare the two methods? This mismatch seems to imply a larger uncertainty than that given by just comparing the means.

This is a good point that deserves discussion. There are two obvious potential causes:

1) The spatial distribution of the inventory is incorrect, such that for all three species it underestimates emissions from south London and/or overestimates emissions from north London.

2) The spatial distribution of emissions in the inventory is broadly correct, but inaccuracy in the model wind field causes the simulated plume to be advected to a slightly more northerly position.

Cause 2) seems to be the more plausible, given that the simulated plume is north of the measured plume for all three species, despite the different source mixes for each species (in particular CH_4). On the other hand, the model wind direction compares fairly well to the measured winds at the aircraft sample locations, and any disagreement between them is actually in the opposite direction (i.e. the

measured winds are more southerly). However, it still seems like too much of a coincidence for all three species to be incorrectly distributed in such a way to generate exactly the same offset in plume position, so we believe inaccuracies in the UKV wind field (prior to reaching the sampling plane) are indeed the likely cause of this mismatch.

The uncertainty associated with the model wind field is one of the main points raised by Referee 2 – please see below for a discussion of this.

Pg 8. Please emphasize throughout the discussion of the comparison that this analysis is for a single flight, and that care should be taken in drawing conclusions about the integrity of the inventory from a single comparison on a single day. Previous authors have shown that when multiple flights are considered, there can be large differences in the calculated flux that are likely due to uncertainties in the top-down flux estimate rather than day-to-day differences in the actual emissions.

Section edited to emphasise this

Pg 8 ln 10-23. I agree that an incorrect spatial pattern in the inventory could explain at least part of the difference. However, I suspect that the choice of background may be more important and be biasing the top-down estimate low. See earlier comments. Does the NAEI include temporal variability and could lack of temporal variability in the NAEI be an explanation for the difference? See earlier comment.

We have edited this section to emphasise the likelihood that temporal variability in emissions (which the NAEI doesn't include) is a likely candidate for this difference. See above for our discussion regarding the choice of background – this can clearly bias the mass balance emission estimates but the flux-dispersion method takes sources within the background footprint into account. Biases associated with inaccuracy in simulated background footprint are discussed at the beginning of Sect. 3.1.2.

Pg 8 ln 24 – 35. It's clear than biogenic CO2 will have an enormous influence on the calculated flux, and that it can bias the CO2 background quite dramatically (see e.g. Turnbull et al 2015, Cambaliza et al 2013). The statement here needs to be much stronger, "treated with caution" is an understatement! It is simply not possible to compare a flux based on total CO2 with an anthropogenic CO2 inventory unless the biogenic component can be accounted for, likely by either having a good biogenic model or being able to separate biogenic and fossil fuel CO2 in the observations (e.g. using 14C or CO). I would say something like "comparison with the NAEI is not appropriate for this dataset".

Edited to make statement stronger

Pg 9 ln 28-34. See earlier comments about choice of background. The same biases occur for this method as for the other method.

This point is discussed in response to the earlier comment

Pg 10 Section 3.2.2. Can you come up with a total emission flux for the flux dispersion method, so that the total flux from each method can be compared more directly? As written, the comparison is between the obs/model ratio for each of the two methods. Thus it can't be determined whether the difference in the ratio occurs because the observed flux rate is different, or the modeled flux rate is different. You argue that the difference is in the modeled flux (actually that you've defined the modelled footprint differently in the two cases). By making a slightly different comparison, this could be argued more strongly.

When we initially started developing the flux-dispersion method, our first approach was to krig the simulated mole fractions and compare total kriged fluxes through the downwind plane. However, we moved away from doing this because by definition any difference in the flux ratios derived in this manner from the overall flux ratios calculated here purely results from the interpolation/extrapolation of sparse data. While the overall flux ratios calculated here are essentially the average of individual transect flux ratios weighted by flux density enhancement, using the kriged flux ratios this average is weighted by a somewhat arbitrary factor relating to the location of the transects relative to each other on the sample plane. In any case, returning to this approach would not tackle the question posed above, because it uses the measured flux calculated with the mass balance method (i.e. the modelled flux is the only value that changes between the two methods).

We have made changes throughout Sect. 3.2 and 3.3 regarding the relationship between the choice of background and the choice of inventory aggregation area (i.e. the "modelled flux") for the mass balance method. Hopefully this explains better the issue with the application of the mass balance method in this study.

Pg 10 section 3.3. This difference in how the footprint is defined is a good candidate for the difference. There are potentially ways to resolve this in the mass balance method. A good start would be to make an estimate of the footprint of the mass balance, rather than assuming that the footprint is an arbitrary metropolitan boundary.

It is not clear to us how an unambiguous definition could be made which separates emissions that contribute to the plume from emissions that contribute to the background (as many source areas contribute to some extent to both). This is discussed at length in response to Referee 2's main point 2).

Pg 11. Conclusions. Please restate the point that the CO2 comparison is invalid because biogenic CO2 is not accounted for. Otherwise the conclusions are very nice.

Point added

Figure 1. Please add a larger scale inset to show where this is in relation to the UK, Ireland, etc. Not all of us are well-versed in English geography!

Inset added

Referee 2:

1) My understanding is that the new approach is merely a combination or trade-off between two traditional approaches : the mass balance approach and the atmospheric transport inversion (Brioude et al. 2013 provide an example of inversion applied to aircraft data around a city). Conceptually, the major difference between this approach and the traditional atmospheric transport inversion is related to the fact that the observed variables to be fitted by rescaling the surface fluxes are fluxes at the measurement locations rather than concentrations. This requires some additional assumptions for the computation of such fluxes, but this enables to account for wind measurements when assimilating the observations. Another difference is that rather than assimilating all local fluxes at the aircraft measurement locations in a Bayesian statistical inversion framework, the method consists here in summarizing them into an average value which is used to rescale the map of surface fluxes. This simplification could lead to a loss of information but it can also help control the inversion behavior. One of the strength of traditional atmospheric inversions and of this new approach is the ability to

extrapolate the information from the sparse measurements by accounting for the atmospheric transport and for the emissions spatial distribution, while the traditional mass balance approach makes coarser extrapolations (here based on a kriging technique).

I think that such a comparison to the atmospheric inversion is worth being discussed since the comparison to the mass balance approach only could lack of hindsight regarding the panel of methods that have been tested to exploit aircraft data. Furthermore, from my point of view, this new approach is closer to the atmospheric transport inversion than to the mass balance approach.

This is a very fair point – in many ways this technique is more similar to previous transport inversion studies than it is to those that use a mass balance method. The focus on mass balancing in the introduction probably reflects the process we went through in developing the technique: we started with a conventional mass balance calculation, but in attempting to resolve the issue of surrounding emission sources each iteration of the technique increasingly revolved around use of the NAME air histories.

We have edited the introduction to better place this method in the context of other approaches used, particularly with regard to atmospheric transport inversions. Flux estimation using a full atmospheric transport inversion is beyond the scope of this study, however a separate study which includes emission estimates for the GAUGE flights derived using a trace gas inversion (employing a hierarchical Bayesian framework) is currently in preparation.

2) One of my main concerns is that by rescaling the total of the NAEI emissions according to measurements whose surface footprint extends well beyond the Greater London area, the new approach does not really inform on the emissions from this area either. Given the distances from the section A-B to London, and as illustrated by Figure 4, the results from this method are driven by emissions from a large part of the South of England that extends to the sea, despite the removal of the "background" concentrations (whose sensitivity to the Western part of the South of England seems much smaller than that of the measurements used to constrain the estimate of emissions according to Figure 4).

Yes, a disadvantage of this new technique is that it is not entirely selective of emissions from a single source area. The key to achieving results relevant to the area of interest is to choose appropriate criteria to define the background, as described in Section 3.1.1. Using the contours in Fig. 4c as a guide it appears reasonable to claim that emissions from the London conurbation represent the most significant influence on the difference between the in-plume and background fluxes, thus supporting the choice of background threshold used here.

This point highlights the key issue with applying the mass balance method in this case, as similarly this is not selective of Greater London emissions but is influenced by emissions from a much wider (but ill-defined) area. In this case the lack of selectivity actively biases the results, an issue which is resolved by applying the new flux-dispersion method. However that both methods suffer from this lack of selectivity was not obvious enough in the original paper – this has now been edited throughout to make it clear.

The computations are conducted in March so that ignoring the natural CO2 fluxes might be fine. But similar computations based on the same aircraft campaigns in spring and summer would be highly hampered by the CO2 uptake upwind and downwind London (not only by the differences between the natural fluxes within the urban part of the measurement footprint vs. within the background footprint that are discussed in section 2.3). While the lack of account for natural CO2 fluxes is mentioned in

section 3.1.2, the major issues raised by these fluxes for spring / summer deserve a discussion, and the topic could deserve some indications in the method sections (in particular in section 2.3) and maybe a coarse look at estimates of the CO2 natural fluxes in the UK.

Yes natural fluxes would be much more significant in summer, and probably cannot be taken as negligible here. We have edited the paper to make a stronger statement regarding the inappropriateness of direct comparison with the NAEI for CO_2 (this point has been particularly highlighted by Referee 1) and have added a paragraph discussing the inferences that can and cannot be drawn regarding natural fluxes by considering the results for CO_2 and CO together (more on this in response to the specific point raised below).

I feel that the manuscript is a bit severe with the mass balance approach by crudely attributing the flux estimate from this method to the Greater London area, and maybe by deriving an estimate of the background concentrations for this approach in a crude way. More cautious interpretations of the flux estimates from this approach are usually made, especially for situations like that of London. I would recommend the authors to comment on the paper by Font et al. (2015) who also made estimates of the emissions from London using aircraft data, and who used FLEXPART simulations to assess the footprint of their measurements. O'Shea et al. (2014) also used NAME to analyze the footprint of their aircraft measurements, and discussed the issue that would be raised by the crude assumption that these measurements would correspond exactly to the greater London area.

It is certainly true to say that aggregating the emissions over the Greater London area is a crude approach. However, it is not clear to us that a more robust method for defining this footprint exists. Font et al. (2015) are able to define a footprint for their measurements using FLEXPART, but their Integrative Mass Boundary Layer method relies on a completely different calculation to the mass balance method used here. They essentially measure the difference between the rate of change in CO_2 concentration within the boundary layer and the rate of entrainment of air from above. This yields a surface flux per unit area, which can then be related to the area the air has travelled over using FLEXPART. In the mass balance approach used here we calculate a bulk flux (i.e. not per unit area) through the downwind sampling plane, relative to our choice of background mole fraction.

The problem here is that, while emissions from certain areas clearly contribute only to the plume, and emissions from certain others contribute only to the background, the majority of the emissions over the air history contribute to both but to different extents. This can clearly be seen in Fig. 4 – summing all of the emissions covered by the air history in Fig. 4b would clearly overestimate the calculated flux (even if they are weighted by residence time, as in Font et al., 2015), as most of these areas are also covered by the air history in Fig. 4a and so contributed to the background. Although this figure relates to the in-plume/background periods for the flux-dispersion method (which is slightly different to the background used in the mass balance calculation taken from the kriged plane), it demonstrates the difficulty in defining any objective criteria for determining the aggregation area. One could subtract the background air history from the in-plume air history and define some threshold value above which a grid square would be included in the aggregated flux total, but the choice of such a threshold would be just as arbitrary as simply using the Greater London boundary. We have added a discussion of this to Section 3.3.

O'Shea et al (2014) took a different approach, using NAME to discard cells from the kriged plane which didn't contain measurements of air that had passed over Greater London. In other words, they stick to aggregating bottom-up emissions over the Greater London area, but try to downscale the kriged flux to represent this area. However, this doesn't solve the fundamental problem with the mass

balance approach, as grid squares with an influence from Greater London can still be influenced by sources surrounding Greater London as well. In the case study presented here, the in-plume measurements were all influenced by London to some degree (as can be seen by comparing Fig. 2 and Fig. 3), so this downscaling approach is not easily applicable.

Therefore, I would be ready to agree that the mass balance approach and its associated type of aircraft measurement tracks is not very well adapted to the monitoring of the emissions from a city surrounded by other cities and productive ecosystems, especially if flight regulations impose measurements to be conducted far downwind. However, I feel that by relying on the same type of measurements and by avoiding to solve for the spatial distribution of the emissions, the new approach may bear the same fundamental limitation which is the lack of ability for isolating the budget of the emissions from the targeted city. In this regard, I think that the conclusions are a bit optimistic.

An additional caveat has been added to the conclusion section on this point.

3) A critical variable in the study is the wind which is used to compute fluxes. Comparisons between measured and modeled (UK Met Office) winds along the transects but also all around the London Greater area could potentially provide some strong insights on the robustness of the transport model, of the estimate of the measurements spatiotemporal footprint and of the estimate of the surface emissions (in particular if biases arise in the comparisons). I feel that it deserves some analysis.

The wind field does indeed have a strong influence on the flux, which is why we have adopted the approach here of taking flux density ratios (rather than concentration ratios). Within the boundary layer the modelled wind speeds are generally higher than the corresponding measurements, both upwind and downwind of Greater London. We account for this by using the modelled wind speed to calculate the simulated flux density, on the assumption that the model overestimation of wind speed will result in a corresponding underestimation in simulated concentration enhancement. We have expanded the discussion of this point in Sect. 3.1.1 and included a figure showing both model and measured wind speeds throughout the flight.

The impact of the overestimated model wind speed on the spatial extent of the footprint is far more subtle and difficult to account for. Given the high bias of the model wind speeds, it is reasonable to assume that the air history likely underestimates the spatial spread of the sample footprint, resulting in the in-plume measurements having a higher simulated sensitivity to emissions from Greater London. This in turn could introduce a low bias into the inventory scale factors.

Any inaccuracy in wind direction across the model wind field also clearly contributes to the uncertainty in this method. Referee 1 points out the mismatch between simulated and measured plume position, which is likely to be a consequence of inaccuracy in the model wind field. This too highlights a potential source of bias, as the air histories for in-plume and background periods simulated by NAME may differ slightly from the actual air histories of the measurements.

The obvious way to investigate the impact of wind field inaccuracy on the uncertainty budget would be to conduct a sensitivity test, where an ensemble of NAME runs are performed driven by a set of met data with perturbed wind fields. This would be an interesting study, but is quite involved from a modelling perspective and goes well beyond the simple use of NAME in this study. We have added a discussion of this potentially significant source of uncertainty in Sect. 3.1.2.

Detailed comments:

- p2l1: explain that "top-down" relates to methods based on atmospheric measurements and models ?

Explicit statement of this added

- p2l5-6: do power plant represent a large fraction of the CO2 emissions in the greater London area ? on the same topic: I had in mind that the city had large power plants in its vicinity that could represent a major share of the emissions in the measurement footprint (http://naei.beis.gov.uk/data/gis-mapping): is it the case ? if yes, it would feed my main concern (2)

Within Great London power plants represent a small fraction of the emissions (<5%). However, power plants in the areas surrounding Greater London do constitute a significant source of emissions, with power plant emissions (including energy from waste facilities) summing to over 25% of the Greater London total. So the referee's main point 2) is valid (see discussion of this above) – it is not possible to isolate the Greater London from surrounding sources using either method presented here, and we have altered the phrasing of the method discussion to clarify this. However, this point also emphasises the advantage of the flux-dispersion method over the mass balance method, as fluxes outside Greater London actively bias the mass balance method results, whereas these surrounding power plant emissions are accounted for in the flux-dispersion method (their presence only acts to reduce the selectivity of the results to the Greater London region).

- p2l15: I am not sure about the meaning of "bulk area flux" here. What would prevent atmospheric inversion to provide such a bulk area flux based on the same data ? see my main point (1)

Yes looking back we agree this is a false distinction – this sentence has been removed anyway in response to main point 1.

- p2l21-33: I feel that the problem of defining the footprint of the estimated flux is presented in an "inverted" way which makes things more complicated than they are. In particular there is no reason to necessarily involve inventories in this problem.

This paragraph has been edited to address the comments of both reviewers. We agree that the emphasis was previously placed too strongly on the comparison with inventory fluxes – the fundamental issue is that, for non-isolated cities, the top-down flux cannot easily be ascribed to a well-defined region. This in turn does make it difficult to compare bottom-up and top-down fluxes using the mass balance method – we have tried to make this narrative clearer in the revised text.

* section 2.2 - this section should provide the duration and the period of the day corresponding to the flight. Maybe I missed it in the following, but the time of the measurements is a critical information that can raise questions regarding the temporal representativity and the robustness of the computations

Requested details added – we agree that this helps to interpret the flux values derived in subsequent sections in the context of the diurnal variability of these fluxes.

- p4l11: I do not understand the end of the sentence ("so as to assess the representativeness ...") in its context

Changed “representativeness” to “accuracy” to clarify what is meant here.

- p4l13: "less than 24 hours" -> 24 hours is large if considering the need to connect the measurements to an emission footprint both in space and in time, and given the strong diurnal variations of the fluxes. Can the statement be more precise based on NAME simulations ?

This has now been made explicit as an average transit time of 20 hours over the British Isles. It is true that diurnal variability in fluxes over this timescale could effectively weight the sensitivity of the derived flux ratios to regions the air passed over during times of peak emission (e.g. rush hour). However, from Fig. 4 it can be seen that the aggregate air histories for in-plume and background sampling begin to converge upwind of London, and certainly by the time the particles leave the domain there is little discernable difference between the distribution of particles for each period. Fluxes in regions that contributed equally to both background and in-plume measurements have no effect on the results, so despite the fact the average transit time over the British Isles is 20 hours, the results are sensitive to fluxes over the much shorter transit time before the two aggregate air histories converge.

* section 3 (beginning) -p5l30: maybe you should already clarify here the fact that NAEI provides annual budgets of the emissions only, while the measurement were made during day time in March, which corresponds to a period of relatively high emissions (this information is limited to the discussions on the CH4 results, and just ignored for CO and CO2 in section 3.1.2). Using constant emissions in the model may also be problematic because the duration of the measurement campaign is about 2.5h, during a period of the day when emissions could be highly variable.

Explicit statement of this added

-p6l1-3: this will be forgotten when discussing the results, while this potentially weakens the confidence in the results from both methods; but this inter-annual change at the national scale may be negligible compared to the seasonal, day-to-day and diurnal variations biasing the comparison between annual budgets in NAEI and the flux estimates for daytime in March (see my comment above)

Point added

* section 3.1.1 -p6l22-23: The sentence (especially "enabling us to specifically assess accuracy. . .") seems to ignore the significant fluxes upwind and downwind London; see my point (2)

Agreed this statement was too strong. This section has been amended in line with point 2) to clarify that, while one can choose a threshold to optimise the selectivity of the results towards the region of interest, there will inevitably be some influence from a wider region.

-p6l31: the latitudinal gradient is not fully accounted for since the background to be removed from local concentrations is taken as a constant value (the average between the north and south backgrounds) rather than as a linear interpolation between the north and south backgrounds; these north and south backgrounds sometimes seem to strongly differ: isn't it an issue (at least as significant as the one raised on p7l1-2) ?

We have repeated the flux-dispersion analysis with an interpolated (rather than averaged) background value used across each transect, and each ratio changed by less than 0.01. We have made reference to this in the paper. See also the discussion of this section in our reply to Referee 1 above.

* section 3.1.2 - p7l17: do the measurements and/or simulations show a significant change of vertical gradients in the concentrations when crossing this BLH ~750m (it does not seem to be the case in Figure 2) ? does the vertical distribution of the concentrations say something about the reliability of the model ?

The vertical distribution of concentrations outside the plume looks similar in the simulated and measured datasets. However, the simulated flux density enhancements are smaller relative to the measured enhancements when sampling above the model boundary layer height. As the simulated dispersion above the model boundary layer height is known to be less accurate we did not include this data.

A plausible explanation for underestimation of the simulated plume magnitude above 750 m would be if the simplified turbulence parameterisation above this height led to the suppression of vertical mixing in the model. In such a case the simulated plume magnitude within the model boundary layer could be consequently overestimated. A full investigation into the impact of turbulence parameterisations on the vertical mixing within the NAME model would require a separate study, but we have added a brief discussion raising this issue as a potential source of bias.

- p7l29: see my main point (2), you need strong assumptions to apply the scaling factors derived for a large part of the South of England to the Greater London area.

Sentence added to re-emphasise the influence of surrounding sources on the results.

- p7l33: I think that this statement is a bit extreme, especially since several investigations could be led to provide insights on the transport uncertainties: the analysis of the wind fields (see my main point (3)), of the 2D vertical structure of the concentration measurements, and, maybe, of the measurements around the Greater London area that are not exploited in this study

Yes fair point – an ensemble NAME run using perturbed met data could be used to quantify some of the uncertainty associated with the dispersion modelling. As stated above, we feel this is beyond the scope of the study here, but we have expanded the discussion of dispersion model uncertainty and highlighted this as a topic for future study.

- p8l20-23 are a bit confusing. I do not really catch how the spatial distribution will be tackled along with the temporal variability.

This section has been reworded to make it clearer. Essentially we are trying to stress the point that more flights would be required to capture temporal emission variability (and reduce the impact of random errors).

- p8l30: the human respiration could also be listed as a source of mismatch ?

We have added explicit reference to this as a component of the higher heterotrophic respiration within the city.

- p8 in a general way: the authors should try to better connect and discuss together the results for CO and CO2: why the scaling factors are so different for these two species ? is it due to the natural CO2 fluxes only ? would not it say something about these natural fluxes ?

We've added a paragraph discussing this. The differences in emission ratios for several key sources makes it a bit of a stretch to attribute all of the difference to the net biospheric flux, but it provides a useful ballpark guide to the potential magnitude of these fluxes.

- p8l31-32: "we can expect them to underestimate" -> shortcut

Changed to "we expect them to underestimate"

* section 3.2.1 -p9l30: "horizontal boundaries" could be rephrased for clarity. Could the definition of the background as the average concentrations over the 15-km boundary sections be too crude for focusing the emission estimate on the Greater London area (is the 15 km distance too short) ? does this background fit well with the background estimated with the flux dispersion method ?

“Horizontal boundaries” has been rephrased. The definition of the background here is crude, but as per the footprint it isn’t clear that a more robust method to define where the plume ends and the background begins exists. Simply eyeballing the data and selecting where the plume ends is a fairly typical approach used in many previous studies. Other studies use some statistical threshold (e.g. 3σ less than the mean mole fraction), but these are in essence equally arbitrary. The average background values used are quite close to those used in the flux-dispersion method – we’ve added these values to the text.

The alternative approach of interpolating (rather than averaging) the background values has been tested. This is discussed in our response to Referee 1 above. We have expanded the discussion on uncertainties associated with the choice of background in Sects. 3.2 and 3.3. See also the response to the final comment below.

-p10l1-3: the discussion goes a bit too fast for me. One could also assume that the upwind concentrations are more suitable to define a background for the measurements downwind London because they would characterize a section across their footprint that is relatively close to the sea (Figs 1 and 4). Discussing the impact of BLH on background concentrations could mean that these background concentrations are mainly driven by fluxes that are relatively close to the measurements. However, the concentrations North and South of the transects A-B are mostly influenced by fluxes North and South of London that are hardly seen by the measurements downwind London, as indicated by Figure 4. In a more general way, I think that the characterization of the "background concentrations" and footprint for the measurements downwind London could be better discussed (see my main point (2)).

We have edited and expanded this section to better discuss the pros and cons of each method for determining the background. Ideally we would define the background so that it represents the mole fractions throughout the downwind plane that would be measured in the absence of any sources from within Greater London. Were we able to calculate this hypothetical value it would solve the issue regarding the definition of aggregation area – if all extraneous sources were accounted for in the background then all enhancements in the downwind plane would only emanate from emissions within Greater London, so aggregating inventory emissions within the area would be reasonable. So, as discussed in our response to Referee 1, the issues raised here with the background definition is essentially another way to express the main issue regarding the inventory aggregation area raised in this paper. We have edited the discussion in Sects. 3.2 and 3.3 to cover this.

Assessing London CO_2, CH_4 and CO emissions using aircraft measurements and dispersion modelling

Joseph R. Pitt[1], Grant Allen[1], Stéphane J.-B. Bauguitte[2], Martin W. Gallagher[1], James D. Lee[3], Will Drysdale[3], Bethany Nelson[3], Alistair J. Manning[4], Paul I. Palmer[5]

[1]School of Earth and Environmental Sciences, University of Manchester, Oxford Road, Manchester, M13 9PL, UK
[2]Facility for Airborne Atmospheric Measurements (FAAM), Cranfield University, Cranfield, MK43 0AL, UK
[3]Wolfson Atmospheric Chemistry Laboratories, Department of Chemistry, University of York, Heslington, York YO10 5DD, UK
[4]Met Office, FitzRoy Road, Exeter, EX1 3PB, UK
[5]School of GeoSciences, University of Edinburgh, Edinburgh, EH9 3FF, UK

Correspondence to: Joseph Pitt (joseph.pitt@manchester.ac.uk)

Abstract. We present a new modelling approach for assessing atmospheric emissions from a city, using an aircraft measurement sampling strategy similar to that employed by previous mass balance studies. Unlike conventional mass balance methods, our approach does not assume that city-scale emissions are confined to a well-defined urban area and that peri-urban emissions are negligible. We apply our new approach to a case study conducted in March 2016, investigating CO, CH_4 and CO_2 emissions from London using aircraft sampling of the downwind plume. For each species, we simulate the flux per unit area that would be observed at the aircraft sampling locations based on emissions from the UK national inventory, transported using a Lagrangian dispersion model. To reconcile this simulation with the measured flux per unit area, assuming the transport model is not biased, we require that inventory values of CO, CH_4, and CO_2 are scaled by 1.00, 0.70, and 1.57, respectively. However, our result for CO_2 ~~must be treated with strong caution as we do not account for the influence of the land and ocean biosphere in this work.~~should not be considered a direct comparison with the inventory which only includes anthropogenic fluxes.

For comparison, we also calculate fluxes using a conventional mass balance approach and compare these to the emissions inventory aggregated over the Greater London area. Using this method we derive much higher inventory scale factors for all three gases, as a direct consequence of ~~neglecting~~ the failure to account for emissions outside the Greater London boundary. That substantially different conclusions are drawn using the conventional mass balance method demonstrates the danger of using this technique for cities whose emissions cannot be separated from significant surrounding sources.

1. Introduction

Over half the people in the world (54%) live in urban areas. This proportion is projected to increase to 66% by 2050 (United Nations, 2014). Consequently, cities are responsible for a large proportion of anthropogenic greenhouse gas (GHG)

emissions. The 2015 UNFCCC Paris Agreement requires signatory states to not only report national GHG emissions, but also to establish and improve independent methods for verifying these reported emissions (UNFCCC, 2015). ~~Therefore the development of t~~Top-down methods that use atmospheric measurements to determine ~~to measure~~ city-scale emissions can assess the accuracy of bottom-up emission inventories and provide crucial information to help improve bottom-up accounting methods.~~has become an increasingly important area of study, aiding both scientific and policy needs.~~

In the UK, spatially and sectorally disaggregated emissions calculated using a bottom-up methodology are given in the National Atmospheric Emissions Inventory (NAEI; Brown et al., 2017). For Greater London, nearly all sources of anthropogenic CO_2 and CO emissions are associated with fuel combustion. For CO_2, the main sources are domestic and commercial combustion and road transport, while emissions from power stations are largely located outside the Greater London administrative boundary. Emissions of CO are comprised of a range of combustion sources, with road transport emissions constituting the largest category. For CH_4 the principal sources are waste treatment and disposal, and leakage during natural gas distribution (in contrast to the UK as a whole, where emissions associated with ruminant livestock dominate). Providing a top-down constraint on these London emissions is important, not only because London represents a large emission source in its own right, but also because it can help inform the assumptions that go into calculating bottom-up emission estimates for these sectors at a national level.

Natural emissions, which are not included in the NAEI, contribute to varying extents for the three species. For CH_4 wetlands are the most significant source of natural emissions within the UK, but wetland fluxes from London and its surrounding areas are negligible compared to anthropogenic emissions. Likewise, in urban areas CO is dominated by anthropogenic sources, although oxidation of biogenic VOCs can contribute, especially during the summer. However, for CO_2 biospheric fluxes not included in the inventory have a significant impact on measured mole fractions downwind of the UK; the impact of these fluxes is discussed further in Sect. 3.1.2.

Formatted: Indent: First line: 1.27 cm

Aircraft mass balance techniques have previously been employed to measure trace gas fluxes from several cities (e.g. Mays et al., 2009; Turnbull et al., 2011; Cambaliza et al., 2014) including London (O'Shea et al., 2014a). ~~These techniques represent a direct method of flux measurement, and unlike atmospheric inversion approaches (e.g. Manning et al., 2011; Bergamaschi et al., 2014; Ganesan et al., 2015) or eddy covariance techniques (e.g. Grimmond et al., 2002; Vogt et al., 2006; Helfter et al., 2011) they can derive a bulk area flux, with a definable uncertainty, using only a few hours of sampling.~~ Typically, horizontal transects are conducted downwind of a city at several altitudes to sample its emitted plume for various trace gas species. These vertically-stacked transects define a 2D plane of sampling downwind of the city. A background mole fraction can be determined by sampling upwind of the city or from downwind measurements outside of the plume. The mass flux of the plume through the 2D plane of sampling is then calculated from the measured mole fraction enhancements (above background) and wind speed.

~~While~~ Using the mass balance method, it is relatively simple to determine the flux for a given species through this 2D sampling plane, relative to a defined background mole fraction. However, attributing this calculated value to a given city,

region or emission source can be far more difficult. In the case of an isolated city, whose surrounding areas can be considered comparatively negligible emission sources, the flux calculated using the mass balance method can be considered representative of the sum total emissions from the city. This enables a direct comparison to be made between top-down and bottom-up emission estimates for the city as a whole. ~~downwind of a city, this value only becomes meaningful if it can be compared to the total city emissions given in the bottom-up inventory. To do so it is necessary to determine a boundary within which inventory emissions can be aggregated to calculate a bulk flux for the city. The choice of this boundary is unproblematic for isolated cities whose surrounding areas can be considered comparatively negligible emission sources.~~ However, for cities such as London, which is surrounded by other non-negligible emission sources for CO_2, CH_4 and CO, it is difficult to associate the flux derived using a conventional mass balance approach with a well-defined spatial region. This in turn makes it hard to directly compare top-down and bottom-up emission estimates, as it is not clear how to ~~determine~~ unambiguously determine the area over which bottom-up emissions should be aggregated to enable comparison with top-down fluxes. ~~calculated using the mass balance method.~~

Another approach involves the use of an atmospheric transport model to represent transport of emitted species from source to measurement site. This enables simulated enhancements to be calculated at the measurement location based on a prescribed emissions map (e.g. from a bottom-up inventory). A range of inverse modelling techniques can then be employed to optimise the emissions map according to the measured enhancements. This is frequently performed at a regional scale using ground-based measurements from long-term monitoring sites (e.g. Manning et al., 2011; Bergamaschi et al., 2015; Ganesan et al., 2015), but it has also been performed using aircraft data to provide the spatial sampling coverage required to estimate city-scale emissions using a few hours of measurement data (e.g. Brioude et al., 2013). One of the key challenges associated with this approach is to control the behaviour of the inversion without over constraining it such that the results only reflect the choice of prior emissions. This is particularly difficult using aircraft data, as for an individual flight the information content of the dataset is limited in comparison to long-term ground-based measurements.

In this study we have developed an alternative method for assessing bottom-up inventory fluxes, using the same aircraft sampling technique ~~described above~~typically employed by mass balance studies. In this new approach we have used the UK Met Office's Lagrangian dispersion model, NAME (Numerical Atmospheric dispersion Modelling Environment), to simulate the transport of the inventory fluxes to the location of the measurements. Rather than optimising an emissions map by using a cost function to reduce model-measurement mismatch (as typically used in previous atmospheric transport inversion studies), we take the simple approach of comparing the average measured and simulated fluxes at the aircraft sample locations and rescale the inventory according to their ratio. This approach effectively constrains the inversion behaviour by removing the freedom to spatially redistribute emissions, while allowing it total freedom to adjust the magnitude of emissions relative to the prior. ~~enables direct comparison between measured and simulated fluxes at the aircraft sample locations, avoiding~~ On the other hand, relative to the mass balance method, we avoid the need to aggregate inventory ~~fluxes~~ emissions over some arbitrarily defined area on the ground in order to compare them to calculated top-down flux values.

In addition to applying our new method, ~~W~~we have also applied the conventional mass balance technique to the same data, and compared the top-down fluxes derived to inventory emissions aggregated over the Greater London administrative area. The two approaches reach significantly different conclusions regarding the accuracy of the emissions inventory; we suggest that this is a consequence of sources outside Greater London that are not adequately accounted for by the background mole fraction calculation, yet nonetheless contribute to the plume measured downwind of the conurbation.

2. Case study details

2.1 Aircraft measurements and calibration

We recorded measurements on board the UK's Facility for Airborne Atmospheric Measurement (FAAM) BAe-146 atmospheric research aircraft (henceforth referred to as the FAAM aircraft). For full details of the aircraft payload see Palmer et al. (2018). Here we describe only those measurements that are relevant to this case study.

Mole fractions of CO_2 and CH_4 were measured using a Fast Greenhouse Gas Analyser (FGGA; Los Gatos Research, USA). Paul et al. (2001) describe the operating principle of the instrument, and O'Shea et al. (2013) provide details on the instrument operational practice and performance on the FAAM aircraft across several campaigns. The FGGA was calibrated hourly in flight, using two calibration gas standards traceable to the WMO-X2007 scale (Tans et al., 2009) and WMO-X2004A scale (an extension of the scale described by Dlugokencky et al., 2005) for CO_2 and CH_4, respectively. For this case study, the certified standards (369.54 ppm CO_2, 1853.8 ppb CH_4 and 456.97 CO_2, 2566.0 ppb CH_4 respectively) spanned the range of measured ambient mole fractions.

We measured a target cylinder containing intermediate mole fraction values approximately half way between these hourly calibrations to quantify any instrument non-linearity or drift. This flight formed part of the wider GAUGE (Greenhouse gAs UK and Global Emissions) campaign, during which we derived average target cylinder measurement offsets of 0.036 ppm for CO_2 and 0.07 ppb for CH_4, relative to the WMO-traceable values, with standard deviations of 0.398 ppm and 2.40 ppb, respectively, for 1 Hz sampling. Each individual target cylinder measurement consisted of a 20 s sample after allowing time for the measurements to reach equilibrium. The standard deviation of these 20 s averaged values was 0.245 ppm for CO_2 and 1.42 ppb for CH_4.

Another source of measurement uncertainty was the impact of water vapour in the sampled air on the retrieved CO_2 and CH_4 mole fractions. This was principally a consequence of mole fraction dilution (i.e. an increase in the total number of molecules per unit volume relative to dry air) and pressure broadening of the spectral absorption lines. The method used to correct for this is described by O'Shea et al. (2013). Using that technique we have derived maximal uncertainties due water vapour of 0.156 ppm for CO_2 and 1.05 ppb for CH_4. Finally, there are also uncertainties associated with the certification of the target cylinder of 0.075 ppm and 0.76 ppb, respectively. Combining all of these uncertainties with the target measurement standard deviations yields nominal total uncertainties for CO_2 and CH_4 of 0.434 ppm and 2.73 ppb at 1 Hz, and 0.300 ppm and 1.93 ppb when averaged over 20 s.

We measured CO mole fractions using vacuum ultraviolet florescence spectroscopy (AL5002, Aerolaser GmbH, Germany). The principle of this system is described by Gerbig et al. (1999), who also evaluate its performance on board an aircraft. Calibration was performed using in-flight measurements of a single gas standard and the background signal at zero CO mole fraction. Gerbig et al. (1999) derive a 1 Hz repeatability of 1.5 ppb (at 100 ppb), and an accuracy of 1.3 ppb ± 2.4% for the 1 Hz measurements.

Details of the meteorological instrumentation on board the FAAM aircraft are provided by Petersen and Renfrew (2009). In summary, we measured temperature with a Rosemount 102AL sensor, with an overall measurement uncertainty of 0.3 K at 95% confidence; we took static pressure measurements from the air data computer, based on measurements from pitot tubes around the fuselage, with an estimated absolute accuracy of 0.5 hPa; we made 3D wind measurements using the a 5-hole probe system described by Brown et al. (1983), with an estimated uncertainty in horizontal wind measurements of < 0.5 m s^{-1}.

2.2 Sampling strategy

On 04 March 2016 we conducted a targeted case study flight to measure CO_2, CH_4 and CO mole fraction enhancements downwind of London, so as to assess the ~~representativeness~~ accuracy of the bottom-up emissions inventory. The flow over the region was consistently westerly, bringing background air from the northern Atlantic with an average travel time over the British Isles of 20 hours~~less than 24 hours~~ (as determined using NAME). As the influence of land-based sources on the recent history of this air mass (upwind of the British Isles) can be assumed to be negligible, we expect that air arriving at the British Isles had relatively homogeneous and well-mixed trace gas composition throughout the boundary layer prior to the influence of local fluxes. Take off was at 08:55 local time (=UTC), with the vertically-stacked transects downwind of London conducted between 11:16 and 13:32 local time. Results from NAME indicate a typical travel time between central London and the downwind sampling plane of ~5 hours, suggesting the majority of the sampled air passed over London between ~06:00 and ~09:00.

Figure 1 shows the flight track from an aerial perspective; between points A and B we flew repeated horizontal transects at various altitudes through a plume of enhanced mole fractions downwind of London emission sources. At the southernmost end of these transects, the constraints of UK airspace forced us to deviate from our desired course perpendicular to the prevailing wind. However, as we sampled the overwhelming majority of the London plume north of this imposed turning point, such that measurements during the deviation to point B represented background (out-of-plume) sampling, we do not expect this deviation to impact on the derived fluxes.

During an initial transect at 1550 m altitude we measured typical uniform free tropospheric background mole fractions for all three gases (CO_2, CH_4 and CO). Following this we descended to 120 m, and subsequently flew 6 transects of increasing altitude, breaking the final transect short to profile up to 1550 m within the observed plume. Figures 2a, 2b and 2c show these transects, coloured by mole fraction for CO_2, CH_4 and CO respectively, projected onto an altitude-latitude plane.

Formatted: Indent: First line: 1.27 cm

2.3 Dispersion model configuration

To determine the air history corresponding to the continuous aircraft sampling we ran the NAME dispersion model in backwards mode, releasing 100 tracer particles at each 1 Hz aircraft measurement location and tracking their motion back in time. NAME was driven by meteorological data from the UK Met Office's UKV model (Tang et al., 2013), which provides hourly data on 70 vertical levels at 1.5 km horizontal resolution over the British Isles. NAME determines particle motion based on the mean wind field (which is determined by interpolating the met data spatially and temporally to the particle location for each time-step) and a parameterisation of unresolved turbulent and mesoscale motions (for details see Jones et al., 2007, and references therein). In this study we used a NAME model time-step of 1 minute. By way of guidance, it is worth noting that although this NAME setup is more computationally intensive than is typically employed, because the release duration was less than 3 hours and the maximum particle travel time before leaving the domain was 37 hours (and less than 24 hours for the majority of particles) the run completion time remained on the order of hours rather than days using the JASMIN scientific computing facility.

To assess the influence of surface fluxes on the sampled air, we used NAME to calculate an air history matrix for each minute of the flight (henceforth referred to as a release period), quantifying the sensitivity of the sampling to the emission in each grid square. To calculate this matrix, we recorded the total time each tracer particle spent in the lowest 100 m above ground level on a 1 km x 1 km horizontal grid (UK National Grid), matching the spatial resolution of the NAEI emissions inventory. The air history matrix D_{ij} was then calculated according to:

$$D_{ij}(t) = \frac{T_{ij}(t)}{d \times n(t)} \quad (1)$$

The indices i and j represent the northing and easting components of the horizontal grid. $T_{ij}(t)$ is the average particle residence time in each grid box for every release period, d is the height of the grid boxes (here 100 m) and $n(t)$ is the average molar density of air (mol_{air} m^{-3}) at the aircraft sampling locations based on UKV data. ~~These individual particle residence times were then averaged over all particles released during each minute of aircraft sampling, giving the average time the sampled air spent in each grid box for every minute of the flight (henceforth referred to as a release period).~~

This air history matrix represents the mole fraction enhancement at the sample locations due to a unit flux in each grid box. ~~A~~By combining this information with the NAEI inventory emissions (F_{ij}) we can calculate a time series of simulated mole fraction enhancements (X) at the aircraft sample locations~~ due to NAEI inventory emissions (F_{ij}) was then calculated according to:~~

$$X(t) = \sum_{i,j} D_{ij}(t) \times F_{ij} \quad (2)$$

~~where the air history matrix D_{ij} represents the mole fraction enhancement at the sample locations due to a unit flux in each grid box, calculated for each release period according to:~~

$$D_{ij}(t) = \frac{T_{ij}(t)}{d \times n(t)} \quad (2)$$

~~The indices i and j in Eqs. (1) and (2) represent the northing and easting components of the horizontal grid. $T_{ij}(t)$ is the average particle residence time in each grid box for every release period, d is the height of the grid boxes (here 100 m) and $n(t)$ is the average molar density of air (mol_{air} m^{-3}) at the aircraft sampling locations based on UKV data.~~ The emissions F_{ij} are given here in units of mol_X m^{-2} s^{-1}, where X represents either CO_2, CH_4 or CO. Figures 2d, 2e and 2f show the equivalent data to Figs. 2a, 2b and 2c, coloured by simulated mole fraction enhancement rather than measured mole fraction.

3. Inventory flux comparisons

In this section we present two approaches to assess the accuracy of the NAEI inventory emission values relative to the measured mole fractions during this case study. The first is a new approach, referred to hereafter as the flux-dispersion method, using the simulated mole fraction enhancements from Sect. 2.3 to derive simulated fluxes through the downwind sampling plane based on inventory emissions, thus enabling comparison with corresponding measured fluxes. The results from this method represent our best assessment of inventory fluxes for this case study.

In Sect. 3.2 we then employ a conventional mass balance method to derive top-down fluxes which are compared to an aggregated NAEI value. We discuss the outcomes of both approaches in Sect. 3.3 and explain how the conventional mass balance approach can lead to spurious conclusions in cases such as this.

It is important to note that the NAEI contains only annually averaged emissions and so does not capture the potentially large temporal variability on diurnal, weekly and seasonal timescales. Clearly this represents a likely source of difference between the top-down results derived from our single flight case study (which represent a snapshot in time) and the inventory; this is discussed further in Sect. 3.1.2. The most recent gridded emissions available in the NAEI at the time of writing were for the year 2015; therefore we have used these 2015 emissions to represent the 2016 values in both approaches. The UK totals (not spatially disaggregated) for 2016 have been released, allowing us to compare these to the 2015 totals. For CO there was a 9.4% reduction in total reported emissions between 2015 and 2016, while for CO_2 there was a reduction of 5.8% and for CH_4 there was an increase of 0.01%. These inter-annual changes are likely to be small in comparison with the variability on shorter timescales mentioned above.

3.1 Flux-dispersion method

3.1.1 Methodology

To make a comparison between the measured and simulated datasets described in Sect. 2 it is first necessary to calculate a background mole fraction for both, so that the mole fraction enhancement due to the London plume can be determined. To determine periods of sampling that were not significantly influenced by the London plume, and therefore can be considered to represent background mole fractions, we again utilised the air history information given by the NAME dispersion

modelling. From the gridded air histories described in Sect. 2.3, we calculated the fraction of $D_{ij}(t)$ that was within the Greater London administrative region for each release period, and defined all release periods where this fraction was less than 0.05% as background periods. This Greater London fraction is shown in Fig. 3, with the background periods coloured red.

In practice there is no sharp distinction between in-plume and background sampling, so any criteria used to separate sampling into these two categories inherently involves some level of human judgement. The key consideration when defining the background for use with this method is to use a threshold that optimises the sensitivity of the results to the region of interest, in this case Greater London. ~~However, it is important to understand that changing this threshold, and therefore selecting different background periods, essentially changes the interpretation of the results as well as the flux values themselves.~~ This is ~~better~~ illustrated by ~~considering~~ Fig. 4, which shows the air history ($D_{ij}(t)$) aggregated over both the background (Fig 4a) and in-plume periods (Fig 4b)~~, which~~ This clearly illustrates that the background criteria used here avoids air histories ~~that have passed over~~with significant influence from the London conurbation.

~~The~~ The comparison between measured and simulated flux discussed later in this section is a comparison between the flux enhancement from the areas sampled in Fig. 4b relative to ~~any~~ the flux enhancement from the air histories sampled in Fig. 4a. Clearly this comparison is not entirely selective of emissions from Greater London, with additional influence from emissions within a wider area (largely upwind and downwind of Greater London). However, given the sampling strategy employed it is not possible to isolate Greater London emissions from other upwind and downwind sources using any technique, and the 0.05% threshold employed represents the best choice to isolate sampling periods with significant Greater London influence. While these extraneous emission sources reduce the selectivity of the results derived using the flux-dispersion method towards Greater London, a key advantage of this technique is that these emissions do not bias the overall results, in contrast to the mass balance method. This is discussed further in Sect. 3.3.~~The justification for choosing the threshold of 0.05% here is essentially that Fig. 4b is dominated by Greater London air histories and Fig. 4a is not, enabling us to specifically assess accuracy of the NAEI for the London conurbation.~~

For both the measured and simulated datasets the mole fraction enhancement due to the London plume is calculated by subtracting the background mole fraction. For each constant-altitude aircraft transect we calculated average background mole fractions to the north and south of the plume separately, for both measured and simulated datasets. We then calculated the mole fraction enhancement, $\Delta X_{London}(t)$, for both datasets using Eq. (3):

$$\Delta X_{London}(t) = X(t) - \frac{\left(\overline{X_{bgd\ N}}(z) + \overline{X_{bgd\ S}}(z)\right)}{2} \quad (3)$$

Here $X(t)$ is the mole fraction time series and $\overline{X_{bgd\ N}}(z)$ and $\overline{X_{bgd\ S}}(z)$ are the average background mole fractions to the north and south of the plume for each transect. The motivation for treating the background in this way is to capture potential latitudinal and vertical gradients in mole fraction. Vertical gradients are accounted for by calculating a separate background value for each transect, while latitudinal gradients are accounted for by the separate calculation of the north and south backgrounds. If a straight average over all background periods was used for each transect, this would be subject to

potential bias in cases where there was more background sampling on one side of the plume than the other. Calculating north and south backgrounds separately as in Eq. (3) mitigates this issue. An alternative method is to interpolate the background values between the north and south edges of the plume, however due to the symmetry of the plume in this case interpolating rather than averaging had a negligible impact on the results (changing the final ratios by less than 1%).

The background values used in Eq. (3) are given in Table 1. It is notable that the simulated CO and CH_4 mole fractions were higher to the north of the plume than to the south, but this gradient was not observed in the measured mole fractions. This could be indicative of an incorrect spatial distribution of emissions within the inventory, which would contradict one of the inherent assumptions made using this method. Alternatively, the higher simulated background to the north of the plume could be counteracted in the measured dataset if this air had a lower initial mole fraction before reaching the British Isles (as the simulated dataset implicitly assumes a uniform background for air entering the domain). The treatment of the background here can mitigate either issue as long as they result in a linear change in mole fraction with latitude. Clearly any non-linear effects have the potential to bias the final results, for example if the air that passed over London had a higher mole fraction before reaching the British Isles than both the background sampling to the north and the south of the plume. While this cannot be entirely ruled out, Fig. 4 shows that (according to NAME) the air sampled during the in-plume periods and the air sampled during the background periods had very similar entry points into the domain, reducing the likelihood that this is a significant source of bias.

The time series of measured and simulated mole fraction enhancements calculated using Eq. (3) are directly comparable quantities. However, the simulated mole fraction enhancements are strongly dependent on the model wind speeds (which have a direct impact on the average particle residence time in Eq. (1)). Any bias in the model wind speeds relative to the measured wind speeds consequently produces a bias in the simulated mole fraction enhancements. Figure 5 shows a comparison of modelled and measured wind speeds throughout the course of the flight. It can be seen that the model tends to overestimate wind speed within the boundary layer, particularly at lower altitudes.~~to reduce the sensitivity of our results to any difference between modelled and measured wind speed,~~

In order to account for the low-biased simulated enhancements resulting from the high-biased model wind speeds, we convert ~~these~~ both measured and simulated mole fraction enhancements into fluxes per unit area in the mean wind direction (i.e. through the downwind sampling plane) before making a comparison between them. To define a representative wind direction, we took the average of the mean UKV model wind direction and the mean measured wind direction during the sampling period. A time series of flux per unit area in this average wind direction, hereafter referred to as the flux density, was then calculated for both measured and simulated datasets using Eq. (4):

$$FD(t) = \Delta X_{London}(t) \times n_{air}(t) \times U_{\perp}(t) \quad (4)$$

The mole fraction enhancement (ΔX_{London}), molar density of air (n_{air}) and wind speed in the mean wind direction ($U_{\perp}$) were calculated independently for the measured and simulated datasets, producing flux densities (FD) in mol m^{-2} s^{-1} that are directly comparable.

3.1.2 Flux-dispersion results

Figure ~~6~~5 shows a comparison between these measured and simulated flux densities as a function of latitude for each plume transect. The lowest transect from Figs. 2 and 3 (~120 m) has ~~also~~ been excluded because no value for $\overline{X_{bgd\ N}}$ was obtained – this was the first transect conducted before the position of the plume had been fully established so its northern extent was not sampled. The top two transects ~~from Figs. 2 and 3~~ have also been excluded here because they are entirely above the average boundary layer height of 759 m used by the NAME dispersion model for this simulation (which is taken directly from the UKV met data). Above this height the parameters used by NAME to describe the turbulent motion of the particles are set to fixed values resulting in poorer representation of particle dispersion. Within the boundary layer these parameters are calculated from the friction velocity and characteristic convective velocity. Therefore the flux densities calculated for transects within the model boundary layer are more accurate than those above it.

We note that the flux density enhancements for the two transects above the model boundary layer are underestimated by the simulation. A possible cause for this would be suppressed vertical mixing in the model as a result of the simplified turbulence parameterisation above this height. A full investigation into the impact of turbulence parameterisation on the vertical mixing within the NAME model would require a separate study, but we note that if the vertical mixing in the model is suppressed this could represent a potential source of bias, leading to larger simulated flux densities within the boundary layer than would in reality be produced by the inventory emissions. ~~The lowest transect (~120 m) has also been excluded because no value for $\overline{X_{bgd\ N}}$ was obtained – this was the first transect conducted before the position of the plume had been fully established so its northern extent was not sampled.~~

A notable feature of the transects shown in Fig. 6 is that the centre of the simulated plume is consistently further north than the centre of the measured plume. This could suggest the spatial distribution of emissions within the inventory is incorrectly weighted towards sources in the north of London. Alternatively, any inaccuracy in the model wind field could lead the simulated plume to be advected to a more northerly position on the sampling plane than the measured plume. The fact that all three species exhibit the same northerly offset of the simulated plume points to the latter explanation, as each species has a different source mix, making it unlikely that they would all exhibit the same spatial bias.

This mismatch in plume position has the potential to impact the results, as it suggests the air histories for in-plume and background periods simulated by NAME may differ slightly from the actual air histories of the measurements. In addition, the high bias of the simulated wind speeds relative to the measurements (shown in Fig. 5) could potentially result in simulated air histories which underestimate the cross-wind spread in the sample footprint. A robust quantification of the uncertainty associated with model wind field inaccuracy would require an ensemble of NAME runs to be performed, driven by met data with perturbed wind fields. Such quantification is beyond the scope of this study, but we note that this is a potentially significant source of uncertainty in the context of the uncertainty ranges calculated below.

Having calculated time series of flux density for the measured and simulated datasets, we then calculated average flux densities for each transect altitude. We also calculated flux densities as an overall average using data from all three

Formatted: Indent: First line: 1.27 cm

transects. These values are given in Table 2~~1~~ for CO_2, CH_4 and CO, along with the ratios between measured and simulated flux densities. These ratios represent the factors by which the NAEI inventory needs to be multiplied in order to reproduce the measured flux densities (assuming there is no bias induced by the NAME transport). Therefore for this case study we conclude that ~~the~~ NAEI emissions ~~for London~~ require scaling by the overall values of 1.00 for CO, 0.70 for CH_4 and 1.57 for CO_2. It is worth stating again that these values are not solely influenced by Greater London emissions, with non-negligible influence from surrounding emission sources, in particular sources immediately upwind and downwind of London. However, through our choice of background (as described in Sect. 3.1.1) we have maximised the sensitivity of the results to emissions from Greater London.

While there are small uncertainties associated with the measured mole fractions (as discussed in Sect. 2.1), the uncertainty in these overall inventory scale factors is expected to be dominated by NAME transport uncertainty~~. This transport uncertainty is fundamentally unknown and a probability distribution of transport error for this case study cannot be determined~~. As discussed above, quantification of the uncertainties associated with the dispersion modelling would require a more involved modelling study using an ensemble of NAME runs. ~~Instead~~ Here we take the range of scale factors across the different transects of 0.90-1.13 for CO, 0.61-0.79 for CH_4 and 1.40-1.77 for CO_2 to give an indication of the total scale factor uncertainty for each species, while noting that there may be additional sources of model transport bias that are not captured by this range.

When considering the implications of the scale factors derived here using the flux-dispersion method, it is important to remember that these correspond to a single flight case study. The variability in emissions on diurnal, weekly and seasonal timescales is potentially large, which places an important caveat on inferences drawn regarding the annual emissions in the NAEI. With this caveat, ~~C~~and considering both the central estimate and uncertainty range ~~for CO~~, we can conclude that the CO measurements taken during this case study are consistent with the emissions given in the NAEI inventory. For CH_4, however, the NAEI emissions yield an overestimate of flux density enhancement for each transect, and overall require downscaling by a factor of 0.70 to agree with observations. This is qualitatively consistent with studies suggesting national NAEI CH_4 emission totals have been too high in previous years (e.g. Manning et al., 2011; Ganesan et al., 2015) but it differs from the most recent top-down verification report (O'Doherty et al., 2017), which finds good agreement between the NAEI CH_4 emission totals and continuous ground-based measurements in recent years.

~~There are multiple possible explanations for the difference between our results for CH_4 emissions and those in the verification report. Our study specifically measures the magnitude of emissions from the London conurbation relative to emissions from surrounding areas. It is possible that although NAEI emission totals agree with long term observations, the spatial distribution of these emissions is not well represented in the inventory, such that the proportion of emissions ascribed to urban areas is too large. Alternatively, as our study represents a single snapshot of emissions taken during a single flight in March 2016, this difference could reflect temporal variability in CH_4 emissions which is not represented in the NAEI.~~ Temporal variability in emissions is an obvious source of difference between our results for CH_4 and those in the verification report. Helfter et al. (2016) used an eddy-covariance technique to determine the diurnal variability of London CH_4 emissions,

and found that emissions increased by a factor of 1.9 (maximum-to-minimum ratio) between the early and late morning. However, in addition to true temporal variability (e.g. rush hour emissions), such techniques are susceptible to the complex nature of urban boundary layer evolution during these transition periods (Halios and Barlow, 2018), which can result in overestimation of diurnal flux variability.

Another possible source of discrepancy with the verification report is the different sample footprint in our study. It is possible that although NAEI emission totals agree with long term observations, the spatial distribution of these emissions is not well represented in the inventory, such that the proportion of emissions ascribed to urban areas is too large. Alternatively, this difference could represent a low-bias in our results associated with inaccuracy in the wind field driving the dispersion modelling (see discussion above). To reduce the impact of both temporal emission variability and random error in the dispersion modelling, allowing for a more robust comparison with the annual NAEI emissions, repeated flights at different times of day, week and year would be required. ~~Ultimately we cannot distinguish between spatial and temporal patterns in emissions using only a single case study. However, future flights could be used in conjunction with the method presented here to build a statistically representative sample of both diurnal and intra-annual flux variability, thus also allowing the spatial distribution of NAEI emissions to be better assessed on an annual timescale.~~

For CO_2 we find that the NAEI would require upscaling in order to be consistent with observations. However, direct comparison with the NAEI is not appropriate for CO_2 because biospheric fluxes, which are not included in the NAEI, represent a significant influence on the measured mole fractions ~~However, this conclusion must be treated with caution as the NAEI does not include biospheric fluxes which are significant for CO_2~~. These biospheric fluxes include uptake due to photosynthesis (gross primary production; GPP), emission from autotrophic respiration and emission from heterotrophic respiration. Net primary production (NPP) is calculated as the difference between photosynthetic uptake and autotrophic respiration. Hardiman et al. (2017) investigated biospheric CO_2 fluxes in Massachusetts and found higher NPP values outside the Boston conurbation. Combined with higher heterotrophic respiration in more populated areas (including human respiration), these higher rural NPP values result in a positive net biospheric flux from urban centres relative to their surrounding areas. As we have not accounted for this net biospheric flux in our simulated flux densities we ~~can~~ expect them to underestimate the measured values, even if the NAEI emissions are entirely accurate.

~~Quantifying~~ Prior quantification of the biospheric impact on the derived scale factor would require the use of an ecosystem model, and is beyond the scope of this study. However, some inferences can be made by considering the different scale factors derived for CO and CO_2, as these species share many of the same combustion sources. Previous studies (e.g. O'Doherty et al., 2013) have used CO measurements as a proxy for anthropogenic CO_2, relying on the assumption that the inventory ratio for CO:CO_2 emissions is correct. In this case, that would imply that the difference in net biospheric flux between the in-plume and background sampling amounted to over half the corresponding difference in anthropogenic flux (comparing the scale factors of 1.00 for CO and 1.57 for CO_2). However, while this comparison can be considered indicative of the potential order of magnitude for net biospheric flux, uncertainty in the inventory CO:CO_2 emission ratio limits our ability to use this method to obtain quantitative information on biospheric fluxes (see Turnbull et al., 2006, for further

Formatted: Indent: First line: 1.27 cm

Formatted: Indent: First line: 1.27 cm

Field Code Changed

discussion on the use of CO:CO_2 ratios for this purpose).~~However the influence of the biosphere must be considered as a key source of uncertainty on our derived CO_2 scale factor.~~

3.2 Conventional mass balance method

3.2.1 Methodology

Detailed descriptions of the mass balance technique in the context of measuring urban GHG emissions are provided by many sources. In general, in the context of bulk area flux measurement, these sources can be categorized into two basic approaches; either the emissions are assumed to be well mixed up to a given height at which they are capped by a temperature inversion (Turnbull et al., 2011; Karion et al., 2013; Smith et al., 2015), or the vertically varying shape of the plume is derived by interpolation between transects flown at multiple altitudes (Mays et al., 2009; Cambaliza et al., 2014; O'Shea et al., 2014a), often using a kriging approach. Figures 2a, 2b and 2c clearly show that the assumptions of the first of these approaches (i.e. well mixed plumes up to a capping height) are not met in this case. We therefore adopt the latter of these approaches, and use kriging to represent the full structure of the plume. This approach necessarily assumes temporal invariance of the plume over the period of sampling: in this case ~2.5 hours.

Following the work of (Mays et al., 2009; Cambaliza et al., 2014; O'Shea et al., 2014a) we derive fluxes using Eq. (5):

$$F = \int_0^{z_{max}} \int_A^B \left(X_{ij} - X_0\right) n_{air}(z)\, U_{\perp ij}\, \mathrm{dx}\, \mathrm{dz} \qquad (5)$$

Here F (mol s^{-1}) is the bulk flux for the emission source, X_{ij} is the kriged mole fraction for a given species, X_0 is the background mole fraction, $n_{air}(z)$ is the molar air density (here derived as a linear function of altitude based on measured values) and $U_{\perp ij}$ is the kriged wind speed perpendicular to the vertical sample plane across which the integral is taken.

Kriging is an interpolation method based on a stochastic Gaussian model, and is described in detail by Kitanidis (1997). It converts samples with sparse spatial coverage into a 2D grid of estimated values, with an associated grid of standard errors for these values. Here we use a modified version of the EasyKrig software (© Dezhang Chu and Woods Hole Ocean Institution) to perform the kriging; again more detail regarding the application of this software with regards to aircraft mass balance flux calculations is given by Mays et al. (2009). More detail regarding the kriging parameters used is included in the supplementary material.

The results from the kriging were output on a 20 x 29 cell grid, with a vertical resolution of 50 m and a horizontal resolution of 5 km respectively, as shown in Fig. 7~~6~~. As the lowest transect was conducted at ~120 m altitude, the structure of the plume below this level was not constrained well by our sampling. Therefore the mole fractions for the lowest 100 m above ground level were taken to be the same as the kriged output for the layer at 100 – 150 m.

~~We determined T~~the background mole fraction X_0 should be chosen to best represent the mole fraction that would be measured downwind of Greater London if there were no emissions within Greater London. We determined this

background for each trace gas by taking the average mole fraction over all cells within 15 km of the ~~horizontal~~ north or south boundary~~ies~~ of the sample plane (i.e. the three columns at each edge of the plane in Fig. 7). This approach follows Mays et al. (2009) in determining the background from measurements in the downwind plane outside of the influence of the plume, and contrasts with the approach used by O'Shea et al. (2014), who instead used measurements upwind of London to determine the background. There are advantages and disadvantages to both methods. Using an upwind background approach such as in O'Shea et al. (2014), one is able to account for sources directly upwind of London in the background measurements, and avoid the influence of extraneous emissions to the north and south of London. However, either multiple transects upwind must be performed to capture the vertical mixing of upwind sources (and hence any vertical gradient in the background) or an assumption must be made that any upwind sources are well-mixed throughout the boundary layer.

A potentially more significant issue with sampling an upwind background prior to the downwind measurements is that, for a morning flight, the boundary layer height can increase significantly during the intervening time. This increase in boundary layer height is associated with entrainment of air from above the boundary layer, which consequently changes the background boundary layer mole fraction. Additionally, in this study we did not sample the same air mass upwind and downwind of London (i.e. we cannot consider this sampling to be in the Lagrangian frame), as the NAME run showed the air took ~9 hours to cover this distance. Therefore the air history for sampling upwind of London may differ significantly from the air history for sampling downwind of London.

These issues with using upwind measurements to calculate a background mole fraction, and the lack of intense sampling within the boundary layer upwind of London in this study, makes a downwind background more appropriate in this case. However, the influence of emission sources to the north and south of London on our calculated background, and the failure to account for emission sources directly upwind and downwind of London, represent key sources of bias in the derived mass balance fluxes below. The impact of this issue is discussed further in Sect. 3.3.

The background mole fractions used were 147.3 ppb for CO, 1941.6 ppb for CH_4 and 409.1 ppm for CO_2, which (despite the difference in background definition) are similar to the values used in the flux-dispersion method (see Table 1). Although we have used these average background values in our main analysis, we have also calculated fluxes using interpolated background values (as recommended by Heimburger et al., 2017) to test the sensitivity of the results to this choice of approach.~~forced to assume that the boundary layer height remains constant, and that there is no exchange of air between the boundary layer and the free troposphere. In comparison, the downwind background method used here only assumes that the boundary layer height and free tropospheric exchange is the same throughout the downwind plane, which represents a less stringent assumption that is likely to hold better in practice.~~

Field Code Changed

3.2.2 Mass balance results

The fluxes calculated using Eq. (5) are given in Table ~~3~~2, along with 1σ uncertainties derived by combining the kriging standard errors with the uncertainty in background mole fraction, taken to be the standard deviation for all background cells used. Also given are the aggregated NAEI emissions for the Greater London administrative area for each species. We have

derived inventory scale factors, in principle analogous to those in Sect. 3.1.2, by taking the ratio of these aggregated NAEI emissions to the flux calculated using Eq. (5) for each species. Using the conventional mass balance method we calculate that the NAEI requires rescaling by factors of 2.27 for CO, 1.22 for CH_4 and 3.08 for CO_2. The differences between these values and those derived in Sect. 3.1.2 are discussed in Sect. 3.3 below. Using interpolated (as opposed to average) background mole fractions slightly increases the calculated fluxes, but in all cases the difference is less than 7%. The NAEI scale factors derived using an interpolated background are calculated to be 2.33 for CO, 1.31 for CH_4 and 3.19 for CO_2.

3.3 Comparing the flux-dispersion and mass balance methods

In Sects. 3.1 and 3.2 two different methods were applied to the same dataset to derive scale factors for the NAEI inventory such that it agrees with aircraft observations. However, the scale factors derived using the flux-dispersion method are significantly lower than those derived using the conventional mass balance method. This is because one of the key elements of the mass balance method, the assumption that a city acts as an isolated emission source surrounded by areas with negligible emissions, is clearly violated in this case. The impact of this is discussed in Sect. 3.2.1 above; the calculated background mole fractions are influenced by extraneous emission sources to the north and south of Greater London, but do not account for sources upwind and downwind of the conurbation. Therefore, the measured mole fraction enhancements above this background do not emanate solely from emissions within Greater London: they are representative of the difference between emissions within the air history of the in-plume measurements and emissions within the air history of the background measurements.

By choosing to aggregate only NAEI emissions within the Greater London administrative area we have ignored ~~any~~ this influence of surrounding emission sources on the calculated mole fraction enhancements, but in the case of London there are significant sources of all three gases outside this administrative area. In particular, ~~L~~leaving upwind and downwind ~~these~~ sources out of the inventory total biases the derived inventory scale factors high for the mass balance method. The influence of these surrounding emission sources could also explain the large inventory upscaling factors derived by O'Shea et al. (2014), who used a mass balance approach to estimate London emissions for a previous case study in 2012. Although an upwind background was used in that study, mitigating the impact of emission sources upwind of London on the derived fluxes, this approach is still susceptible to biases associated with emissions downwind and to the north and south of London, which all contribute to the measured mole fraction enhancements in that study.

Clearly one could choose a different region over which to aggregate NAEI emissions, but for a non-isolated source such a choice is inherently arbitrary and yet has a direct impact on the derived inventory scale factor. Just as ignoring sources which contribute to the plume resulted here in high-biased scale factors, aggregating over too wide an area would introduce a low bias, as sources which actually contribute predominantly to the measured background are instead assumed to contribute to the plume. Emissions from many source areas contribute to some extent to both the in-plume and background measurements, making it unclear whether or not to include these in the aggregated inventory total.

Formatted: Indent: First line: 1.27 cm

It is worth noting that Lagrangian mass balance techniques (e.g. O'Shea et al., 2014b) and Integrative Mass Boundary Layer techniques (e.g. Font et al., 2015) frequently use dispersion model air histories to define the measured flux footprint. However, the methodology employed by these techniques, i.e. balancing the change in species concentration within a column of air against the various fluxes per unit area into and out of the column, lends itself more naturally to such a flux footprint definition, as it does not rely on defining a spatially separate period of background sampling. Using the downwind mass balance approach here, we cannot simply attribute the derived mass balance flux to the area given by the NAME air history for in-plume sampling, as the emissions from much of this area contribute to the background sampling as well (this can clearly be seen from Fig. 4). The techniques referred to above provide alternative methods for quantifying emissions from urban areas, although they generally have to assume emissions are well-mixed throughout the boundary layer (which is not always the case over urban areas), and obtaining permission to fly at low level directly over a city can present an additional logistical issue.

~~T~~In summary, the impact of the arbitrary choice of inventory aggregation area on the conclusions drawn using the mass balance method demonstrates the inappropriateness of the downwind mass balance technique for non-isolated emission sources. Instead, the flux-dispersion method provides a good alternative in these cases, as it is not subject to such biases.

5. Conclusions

Aircraft mass balance techniques are an effective way of determining emissions from isolated sources, but they require surrounding areas to be negligible emission sources in order to yield robust results. This is a well-known assumption associated with these methods. However in the absence of alternative techniques using the same sample dataset against which the mass balance results can be compared, one is forced either to simply state this assumption as a caveat or to abandon the effort entirely.

In this study we have developed an alternative technique using a Lagrangian dispersion model to quantify the transport of inventory emissions to the aircraft sample locations, so that a direct comparison of flux per unit area can be made at the measurement locations. In contrast to the conventional mass balance technique, this method does not require cities to be isolated from surrounding emission sources, rendering it more appropriate in many cases. We have demonstrated this new technique by applying it to a single-flight case study measuring London emissions, which yielded inventory scale factors of 1.00 (0.90-1.13) for CO, 0.70 (0.61-0.79) for CH_4 and 1.57 (1.40-1.77) for CO_2. These values represent the factors by which the inventory emissions need to be multiplied to agree with the aircraft measurements, although the absence of biospheric fluxes in the inventory means direct comparison with the CO_2 measurements is not appropriate. Using a mass balance approach we derived significantly higher values (2.27, 1.22 and 3.08 respectively), which we conclude are biased as a consequence of ~~choosing an inventory aggregation area that ignores~~ significant sources outside the Greater London administrative region, which are neither adequately captured by the background mole fraction calculation or easy to account for in the choice of inventory aggregation area. The magnitude of this bias demonstrates how employing a mass balance

method for a non-isolated source can lead to highly misleading conclusions regarding the accuracy of the emissions inventory under study.

It is important to emphasise that the inventory scale factors derived here represent the results from a single case study and therefore are not necessarily representative of the annual timescale of the NAEI emissions. In order to better validate the inventory on this timescale repeated flights following a similar sampling strategy are required. The limited spatial selectivity of the flux-dispersion technique represents another caveat on the results from a single flight, as the derived flux ratios are not only sensitive to emissions from the London conurbation but also to emissions from a fairly wide area surrounding it. Repeated flights should therefore be designed to incorporate sampling under different prevailing wind directions, so that the systematic impact of extraneous sources on the overall results is minimised. Using the flux-dispersion method developed here in combination with representative aircraft sampling on an annual timescale could provide a robust assessment of inventory fluxes at the city scale in the case of non-isolated sources for which the mass balance technique is not appropriate.

Author contributions

Conceptualization, J. P., G. A., J. L.; Data curation; J. P., S. B.; Formal analysis; J. P.; Funding acquisition; G. A., M. G., J. L., A. M., P. P.; Investigation; J. P., G. A., S. B., J. L., W. D., B. N.; Methodology; J. P., G. A.; Software, A. M.; Supervision, G. A., M. G., J. L., P. P.; Writing – original draft, J. P.; Writing – review & editing, J. P. G. A., P. P.

Acknowledgements

The authors would like to thank everyone at FAAM, Airtask and Avalon Aero who helped with the collection and processing of the aircraft data used here. We acknowledge use of the NAME atmospheric dispersion model and associated NWP meteorological data sets made available to us by the Met Office. We acknowledge the significant storage resources and analysis facilities made available to us on JASMIN by STFC CEDA along with the corresponding support teams. J. R. Pitt received a NERC CASE studentship in partnership with FAAM, grant number NE/L501/591/1, supervised by G. Allen. This work was supported by the GAUGE (Greenhouse gAs UK and Global Emissions) NERC project, grant numbers NE/K002449/1 and NE/K00221X/1.

References

Bergamaschi, P., Corazza, M., Karstens, U., Athanassiadou, M., Thompson, R. L., Pison, I., Manning, A. J., Bousquet, P., Segers, A., Vermeulen, A. T., Janssens-Maenhout, G., Schmidt, M., Ramonet, M., Meinhardt, F., Aalto, T., Haszpra, L., Moncrieff, J., Popa, M. E., Lowry, D., Steinbacher, M., Jordan, A., O'Doherty, S., Piacentino, S. and Dlugokencky, E.: Top-

down estimates of European CH_4 and N_2O emissions based on four different inverse models, Atmos. Chem. Phys., 15, 715–736, doi:10.5194/acp-15-715-2015, 2015.

Brioude, J., Angevine, W. M., Ahmadov, R., Kim, S.-W., Evan, S., McKeen, S. A., Hsie, E.-Y., Frost, G. J., Neuman, J. A., Pollack, I. B., Peischl, J., Ryerson, T. B., Holloway, J., Brown, S. S., Nowak, J. B., Roberts, J. M., Wofsy, S. C., Santoni, G. W., Oda, T. and Trainer, M.: Top-down estimate of surface flux in the Los Angeles Basin using a mesoscale inverse modeling technique: assessing anthropogenic emissions of CO, NOx and CO2 and their impacts, Atmos. Chem. Phys., 13, 3661–3677, doi:10.5194/acp-13-3661-2013, 2013.

Brown, E. N., Friehe, C. A. and Lenschow, D. H.: The Use of Pressure Fluctuations on the Nose of an Aircraft for Measuring Air Motion, J. Clim. Appl. Meteorol., 22, 171–180, doi:10.1175/1520-0450(1983)022<0171:TUOPFO>2.0.CO;2, 1983.

Brown, P., Broomfield, M., Cardenas, L., Choudrie, S., Kilroy, E., Jones, L., Passant, N., Thomson, A., Wakeling, D., Buys, G., Forden, S., Gilhespy, S., Glendining, M., Gluckman, R., Henshall, P., Hobson, M., MacCarthy, J., Malcolm, H., Manning, A., Matthews, R., Milne, A., Misselbrook, T., Moxley, J., Murrells, T., Salisbury, E., Sussams, J., Thistlethwaite, G., Walker, C. and Webb, N.: UK Greenhouse Gas Inventory, 1990 to 2015, Ricardo Energy & Environment, Harwell, UK, 2017.

Cambaliza, M. O. L., Shepson, P. B., Caulton, D. R., Stirm, B., Samarov, D., Gurney, K. R., Turnbull, J., Davis, K. J., Possolo, A., Karion, A., Sweeney, C., Moser, B., Hendricks, A., Lauvaux, T., Mays, K., Whetstone, J., Huang, J., Razlivanov, I., Miles, N. L. and Richardson, S. J.: Assessment of uncertainties of an aircraft-based mass balance approach for quantifying urban greenhouse gas emissions, Atmos. Chem. Phys., 14, 9029–9050, doi:10.5194/acp-14-9029-2014, 2014.

Dlugokencky, E. J., Myers, R. C., Lang, P. M., Masarie, K. A., Crotwell, A. M., Thoning, K. W., Hall, B. D., Elkins, J. W. and Steele, L. P.: Conversion of NOAA atmospheric dry air CH4 mole fractions to a gravimetrically prepared standard scale, J. Geophys. Res., 110, D18306, doi:10.1029/2005JD006035, 2005.

Font, A., Grimmond, C. S. B., Kotthaus, S., Morguí, J.-A., Stockdale, C., O'Connor, E., Priestman, M. and Barratt, B.: Daytime CO2 urban surface fluxes from airborne measurements, eddy-covariance observations and emissions inventory in Greater London, Environ. Pollut., 196, 98–106, doi:10.1016/j.envpol.2014.10.001, 2015.

Ganesan, A. L., Manning, A. J., Grant, A., Young, D., Oram, D. . E., Sturges, W. T., Moncrieff, J. B. and O'Doherty, S.: Quantifying methane and nitrous oxide emissions from the UK and Ireland using a national-scale monitoring network, Atmos. Chem. Phys., 15, 6393–6406, doi:10.5194/acp-15-6393-2015, 2015.

Gerbig, C., Schmitgen, S., Kley, D., Volz-Thomas, A., Dewey, K. and Haaks, D.: An improved fast-response vacuum-UV resonance fluorescence CO instrument, J. Geophys. Res. Atmos., 104, 1699–1704, doi:10.1029/1998JD100031, 1999.

Halios, C. H. and Barlow, J. F.: Observations of the Morning Development of the Urban Boundary Layer Over London, UK, Taken During the ACTUAL Project, Boundary-Layer Meteorol., 166, 395–422, doi:10.1007/s10546-017-0300-z, 2018.

Hardiman, B. S., Wang, J. A., Hutyra, L. R., Gately, C. K., Getson, J. M. and Friedl, M. A.: Accounting for urban biogenic fluxes in regional carbon budgets, Sci. Total Environ., 592, 366–372, doi:10.1016/j.scitotenv.2017.03.028, 2017.

Heimburger, A. M. F., Harvey, R. M., Shepson, P. B., Stirm, B. H., Gore, C., Turnbull, J., Cambaliza, M. O. L., Salmon, O. E., Kerlo, A.-E. M., Lavoie, T. N., Davis, K. J., Lauvaux, T., Karion, A., Sweeney, C., Brewer, W. A., Hardesty, R. M. and Gurney, K. R.: Assessing the optimized precision of the aircraft mass balance method for measurement of urban greenhouse gas emission rates through averaging, Elem Sci Anth, 5, 26, doi:10.1525/elementa.134, 2017.

Helfter, C., Tremper, A. H., Halios, C. H., Kotthaus, S., Bjorkegren, A., Grimmond, C. S. B., Barlow, J. F. and Nemitz, E.: Spatial and temporal variability of urban fluxes of methane, carbon monoxide and carbon dioxide above London, UK, Atmos. Chem. Phys., 16, 10543–10557, doi:10.5194/acp-16-10543-2016, 2016.

Jones, A., Thomson, D., Matthew Hort and Devenish, B.: The U.K. Met Office's Next-Generation Atmospheric Dispersion Model, NAME III, in: Air Pollution Modeling and Its Application XVII, edited by C. Borrego and A.-L. Norman, 580–589, Springer US, New York, NY, USA., 2007.

Karion, A., Sweeney, C., Pétron, G., Frost, G., Hardesty, R. M., Kofler, J., Miller, B. R., Newberger, T., Wolter, S., Banta, R., Brewer, A., Dlugokencky, E., Lang, P., Montzka, S. A., Schnell, R., Tans, P., Trainer, M., Zamora, R. and Conley, S.: Methane emissions estimate from airborne measurements over a western United States natural gas field, Geophys. Res. Lett., 40, 4393–4397, doi:10.1002/grl.50811, 2013.

Kitanidis, P. K.: Introduction to Geostatistics: Applications in Hydrogeology, Cambridge University Press, Cambridge, U.K., 1997.

Manning, A. J., O'Doherty, S., Jones, A. R., Simmonds, P. G. and Derwent, R. G.: Estimating UK methane and nitrous oxide emissions from 1990 to 2007 using an inversion modeling approach, J. Geophys. Res., 116, D02305, doi:10.1029/2010JD014763, 2011.

Mays, K. L., Shepson, P. B., Stirm, B. H., Karion, A., Sweeney, C. and Gurney, K. R.: Aircraft-based measurements of the carbon footprint of Indianapolis., Environ. Sci. Technol., 43, 7816–7823, doi:10.1021/es901326b, 2009.

O'Doherty, S., Grant, A., Manning, A. J., Derwent, R. G., Simmonds, P. and Young, D.: Long-term atmospheric measurement and interpretation (of radiatively active trace gases), Report to DECC (Department of Energy and Climate Change), Contract number: GA0201, 2013.

O'Doherty, S., Stanley, K., Rigby, M., Stavert, A., Manning, A., Redington, A., Simmonds, P., Young, D. and Sturges, B.: Long-term atmospheric measurement and interpretation (of radiatively active trace gases), Report to BEIS (Department for Business, Energy and Industrial Strategy), Contract number: GA0201, 2017.

O'Shea, S. J., Bauguitte, S. J.-B., Gallagher, M. W., Lowry, D. and Percival, C. J.: Development of a cavity-enhanced absorption spectrometer for airborne measurements of CH4 and CO2, Atmos. Meas. Tech., 6, 1095–1109, doi:10.5194/amt-6-1095-2013, 2013.

O'Shea, S. J., Allen, G., Fleming, Z. L., Bauguitte, S. J.-B., Percival, C. J., Gallagher, M. W., Lee, J., Helfter, C. and Nemitz, E.: Area fluxes of carbon dioxide, methane, and carbon monoxide derived from airborne measurements around Greater London: A case study during summer 2012, J. Geophys. Res. Atmos., 119, 4940–4952, doi:10.1002/2013JD021269, 2014a.

O'Shea, S. J., Allen, G., Gallagher, M. W., Bower, K., Illingworth, S. M., Muller, J. B. A., Jones, B. T., Percival, C. J.,

Bauguitte, S. J.-B., Cain, M., Warwick, N., Quiquet, A., Skiba, U., Drewer, J., Dinsmore, K., Nisbet, E. G., Lowry, D., Fisher, R. E., France, J. L., Aurela, M., Lohila, A., Hayman, G., George, C., Clark, D. B., Manning, A. J., Friend, A. D. and Pyle, J.: Methane and carbon dioxide fluxes and their regional scalability for the European Arctic wetlands during the MAMM project in summer 2012, Atmos. Chem. Phys., 14, 13159–13174, doi:10.5194/acp-14-13159-2014, 2014b.

Palmer, P. I., O'Doherty, S., Allen, G., Bower, K., Bösch, H., Chipperfield, M. P., Connors, S., Dhomse, S., Feng, L., Finch, D. P., Gallagher, M. W., Gloor, E., Gonzi, S., Harris, N. R. P., Helfter, C., Humpage, N., Kerridge, B., Knappett, D., Jones, R. L., Le Breton, M., Lunt, M. F., Manning, A. J., Matthiesen, S., Muller, J. B. A., Mullinger, N., Nemitz, E., O'Shea, S., Parker, R. J., Percival, C. J., Pitt, J., Riddick, S. N., Rigby, M., Sembhi, H., Siddans, R., Skelton, R. L., Smith, P., Sonderfeld, H., Stanley, K., Stavert, A. R., Wenger, A., White, E., Wilson, C. and Young, D.: A measurement-based verification framework for UK greenhouse gas emissions: an overview of the Greenhouse gAs Uk and Global Emissions (GAUGE) project, Atmos. Chem. Phys., 18, 11753–11777, doi:10.5194/acp-18-11753-2018, 2018.

Paul, J. B., Lapson, L. and Anderson, J. G.: Ultrasensitive absorption spectroscopy with a high-finesse optical cavity and off-axis alignment., Appl. Opt., 40, 4904–4910, 2001.

Petersen, G. N. and Renfrew, I. A.: Aircraft-based observations of air-sea fluxes over Denmark Strait and the Irminger Sea during high wind speed conditions, Q. J. R. Meteorol. Soc., 135, 2030–2045, doi:10.1002/qj.355, 2009.

Smith, M. L., Kort, E. A., Karion, A., Sweeney, C., Herndon, S. C. and Yacovitch, T. I.: Airborne Ethane Observations in the Barnett Shale: Quantification of Ethane Flux and Attribution of Methane Emissions, Environ. Sci. Technol., 49, 8158–8166, doi:10.1021/acs.est.5b00219, 2015.

Tang, Y., Lean, H. W. and Bornemann, J.: The benefits of the Met Office variable resolution NWP model for forecasting convection, Meteorol. Appl., 20, 417–426, doi:10.1002/met.1300, 2013.

Tans, P., Zhao, C. and Kitzis, D.: The WMO Mole Fraction Scales for CO2 and other greenhouse gases, and uncertainty of the atmospheric measurements, in: 15th WMO/IAEA Meeting of Experts on Carbon Dioxide, Other Greenhouse Gases, and Related Tracer Measurement Techniques, Jena, Germany., 2009.

Turnbull, J. C., Miller, J. B., Lehman, S. J., Tans, P. P., Sparks, R. J. and Southon, J.: Comparison of 14CO2 , CO, and SF6 as tracers for recently added fossil fuel CO2 in the atmosphere and implications for biological CO2 exchange, Geophys. Res. Lett., 33, L01817, doi:10.1029/2005GL024213, 2006.

Turnbull, J. C., Karion, A., Fischer, M. L., Faloona, I., Guilderson, T., Lehman, S. J., Miller, B. R., Miller, J. B., Montzka, S., Sherwood, T., Saripalli, S., Sweeney, C. and Tans, P. P.: Assessment of fossil fuel carbon dioxide and other anthropogenic trace gas emissions from airborne measurements over Sacramento, California in spring 2009, Atmos. Chem. Phys., 11, 705–721, doi:10.5194/acp-11-705-2011, 2011.

UNFCCC: Adoption of the Paris Agreement, in: Conference of the Parties: Twenty-first session, Paris, France., 2015.

United Nations: World Urbanization Prospects: The 2014 Revision, Highlights (ST/ESA/SER.A/352), United Nations, Department of Economic and Social Affairs, Population Division, 2014.

Dataset	Alt (m)	CO (ppb)		CH_4 (ppb)		CO_2 (ppm)	
		South	North	South	North	South	North
Measured	287	148.4	147.4	1943.1	1941.8	409.6	409.3
	460	149.2	145.3	1942.5	1938.4	409.3	408.9
	575	150.2	149.9	1943.6	1943.3	409.2	409.3
Simulated	287	3.7	4.5	6.8	11.7	0.5	0.6
	460	3.7	3.8	7.0	10.7	0.5	0.5
	575	3.7	4.3	6.9	12.7	0.5	0.5

Table 1: Background mole fractions for each species to the north and the south of the London plume, calculated using the flux-dispersion method.

Formatted: Heading 6

	CO			CH_4			CO_2		
	Mean flux density ($\mu mol\ m^{-2}\ s^{-1}$)			Mean flux density ($\mu mol\ m^{-2}\ s^{-1}$)			Mean flux density ($\mu mol\ m^{-2}\ s^{-1}$)		
Alt (m)	Measured	Simulated	Ratio	Measured	Simulated	Ratio	Measured	Simulated	Ratio
287	1.80	1.80	1.00	2.73	3.46	0.79	541	348	1.56
460	1.65	1.84	0.90	2.12	3.50	0.61	468	334	1.40
575	1.94	1.71	1.13	1.96	2.85	0.69	556	314	1.77
Overall	1.79	1.79	**1.00**	2.28	3.28	**0.70**	521	332	**1.57**

Table 2~~1~~: Mean flux densities calculated using the flux-dispersion method, given for each transect and taken over all three transects. The ratios between measured and simulated flux densities are all given.

	CO		CH_4		CO_2	
	Mean	1σ	Mean	1σ	Mean	1σ
Flux (kmol s^{-1})	0.178	0.006	0.182	0.009	44.7	1.2
NAEI emissions (kmol s^{-1})	0.079	N/A	0.149	N/A	14.5	N/A
Ratio	**2.27**	0.07	**1.22**	0.06	**3.08**	0.08

Table 32: Bulk fluxes calculated using a conventional mass balance technique and corresponding NAEI emissions, aggregated over the Greater London administrative region. The ratio of mass balance flux to NAEI emission is also given.

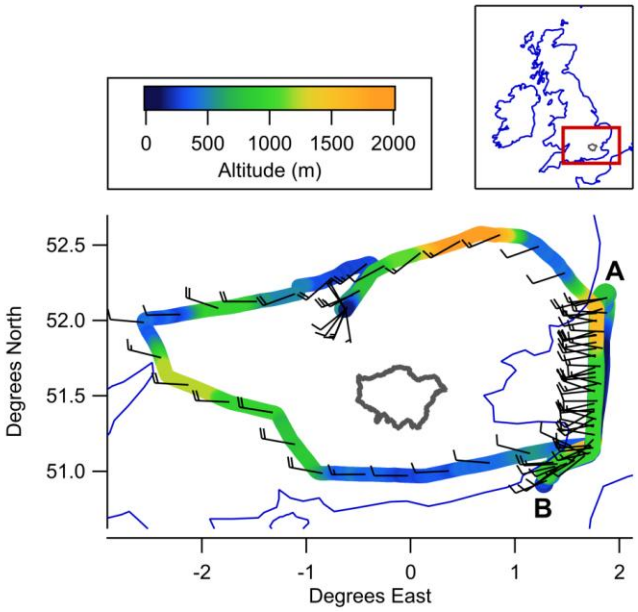

Figure 1: Aircraft flight track on 04 March 2016, coloured by altitude. Wind barbs are used to represent wind speed and direction, averaged over 5 minutes, using the convention where each full wind barb represents a wind speed of 10 knots. The border of the Greater London administrative region is shown in grey for reference.

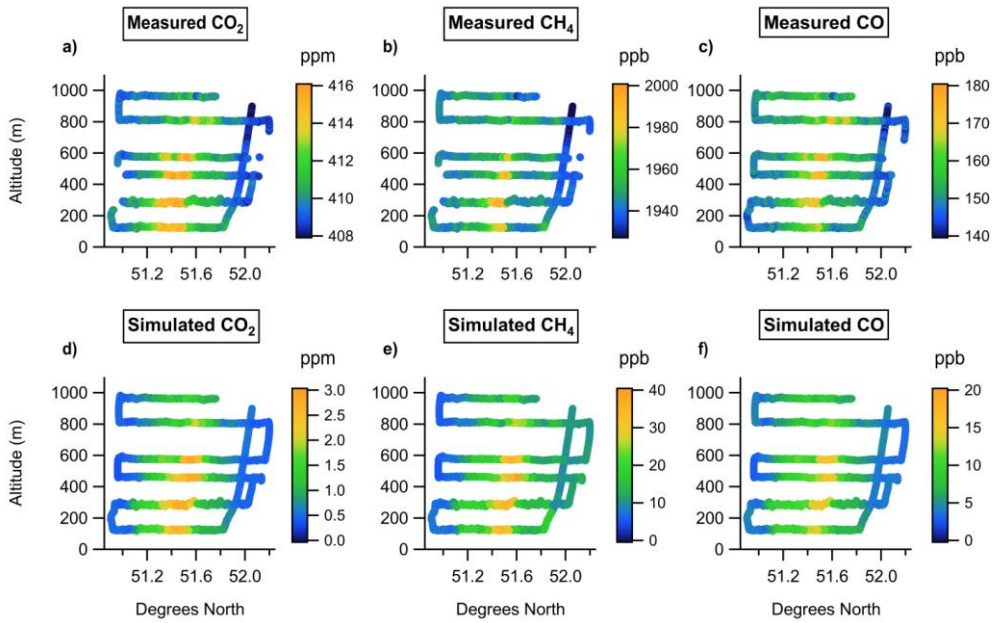

Figure 2: Altitude-latitude projections of measured mole fraction (a – c) and simulated (d – f) mole fraction enhancement downwind of London for each species.

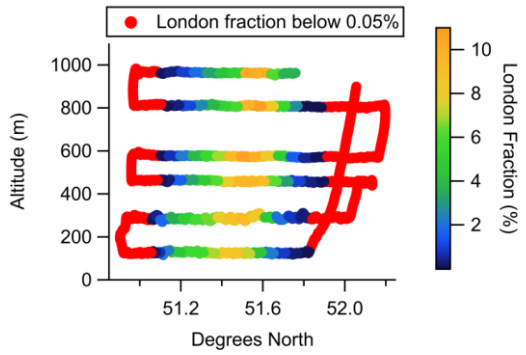

Figure 3: Altitude-latitude projection showing the influence of London on the downwind sampling, as determined from the NAME air histories. The colour scale represents the fraction of aggregated NAME air history D_{ij} within the Greater London administrative region for each NAME release period. Background periods, where the London fraction is less than 0.05%, are shown in red.

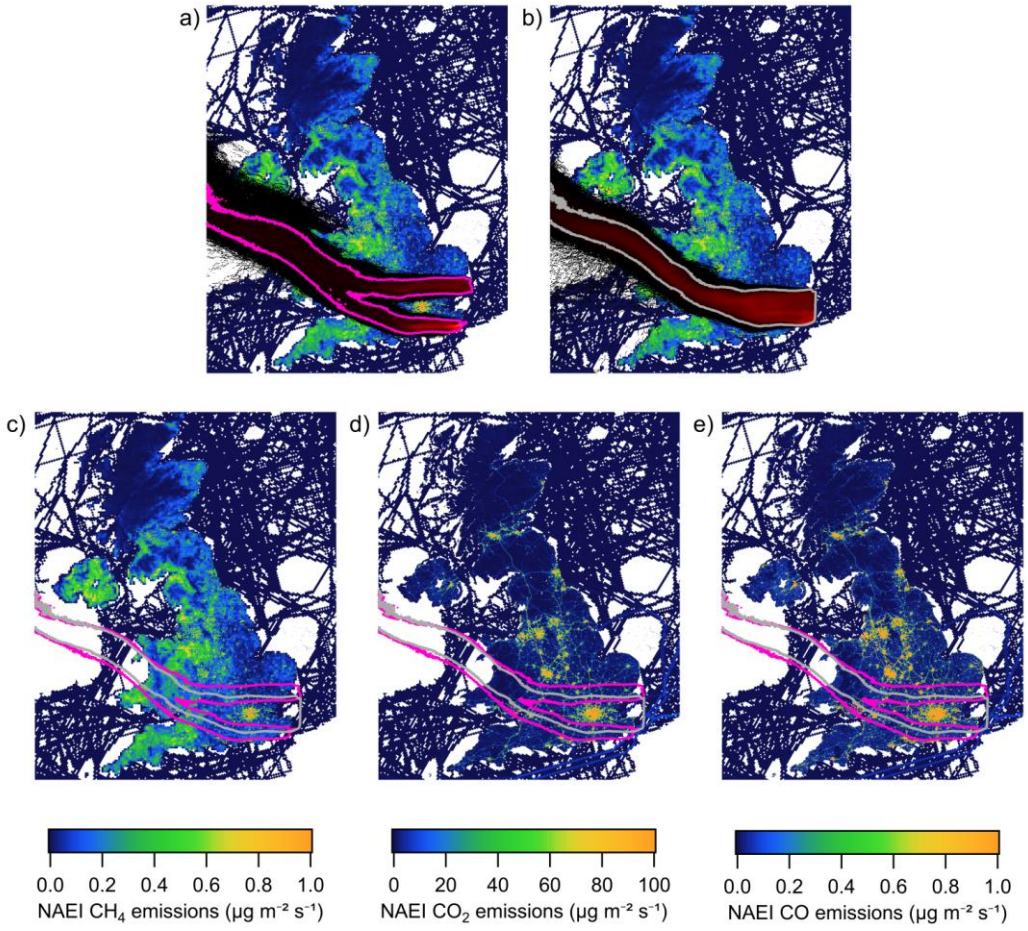

Figure 4: NAME air histories aggregated over (a) background sampling periods, and b) in-plume sampling periods, overlaid on an NAEI emissions inventory map for CH_4 (shown using a saturated colour scale). Both air histories have been normalised such that they sum to 1, with grey and pink contours shown in each plot surrounding the vast majority (99.9995%) of sample influence. These contours are included in panel c), d) and e) to provide a better visual comparison between the two aggregate air histories in the context of the inventory emissions for CH_4, CO_2 and CO respectively.

Formatted: Font: 9 pt, Bold

Formatted: Justified, Line spacing: 1.5 lines

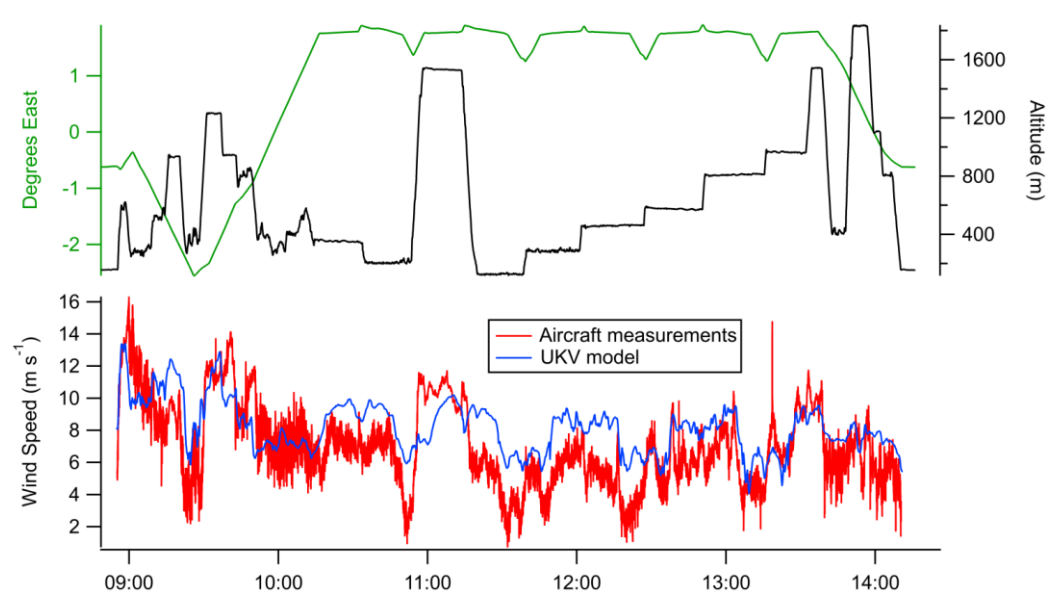

Figure 5: Comparison between wind speeds measured by the aircraft and the corresponding wind speeds at the aircraft location from the UKV model. It can be seen that the model generally overestimates wind speed within the boundary layer.

Formatted: Font: 9 pt, Bold

Formatted: Heading 6

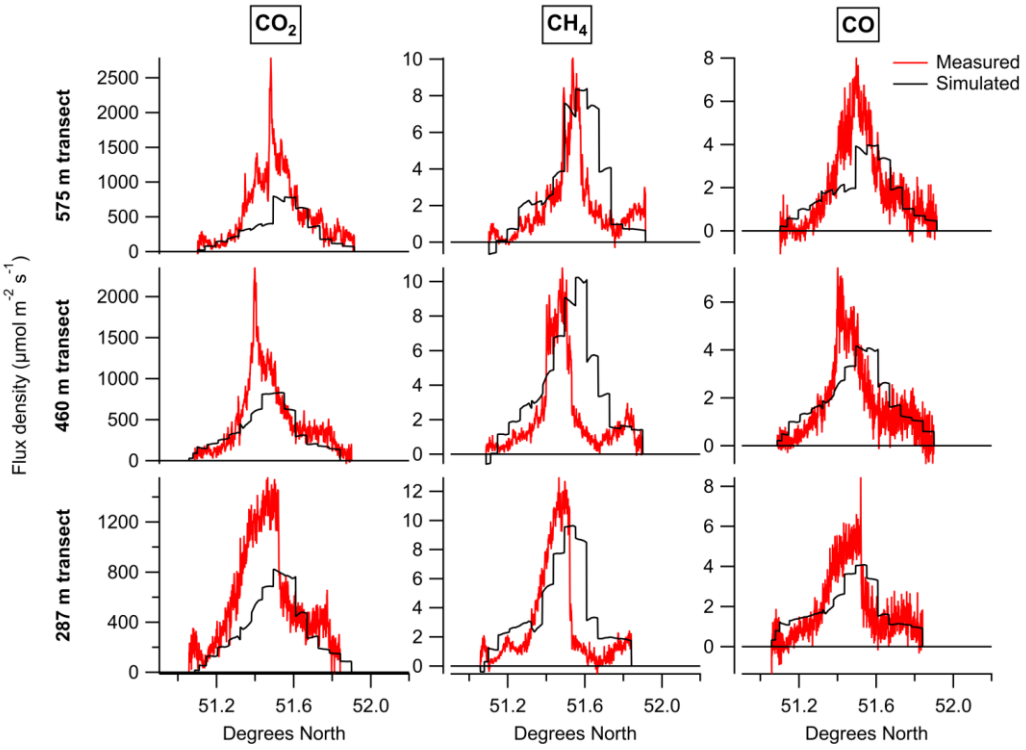

Figure 65: Measured and simulated flux densities for CO_2, CH_4 and CO, given for the three transects (287 m, 460 m and 575 m) used to assess inventory accuracy.

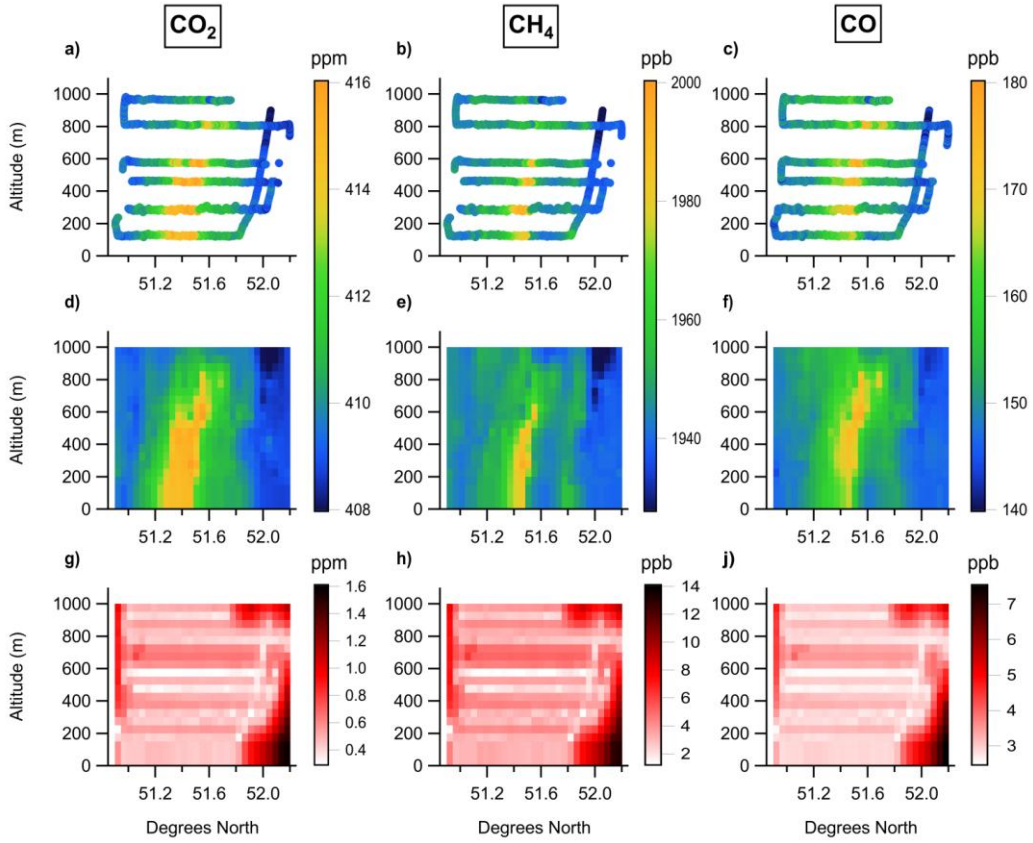

Figure 76: Altitude-latitude projections of: a) – c) measured data, d) – f) kriged data, g) – j) kriging standard error, for CO_2, CH_4 and CO respectively.