

Anonymous Referee #1

General Comments:

Using five atmospheric CO₂ monitoring sites, this paper presents top-down (inversion) based CO₂ flux estimations for a region around Paris. In this study, Chimere transport model in combination with different flux datasets (both anthropogenic and biospheric) is used to simulate atmospheric mole fractions of CO₂. The inversion is based on Bayesian approach and is performed independently in two different set-ups: one set-up uses the mole fraction measurements and the other uses concentration gradients as measurement vector (y). The study further investigates on the usability of Eiffel tower measurements in these inversion set-ups to generate meaningful flux estimations. Overall, this paper holds valuable information on this area of research, although some analysis needs further clarification. The major issues are mentioned below.

We agree with this summary, and thank the reviewer for his work and suggestions

Major Comments:

My major concern is that the paper is often too far away from being succinct and clear, which distracts readers a lot. The problem is not only with the inclusion of irrelevant information (for e.g. figure legend explanation inside the manuscript text that could be done (or already done) in respective figure caption), but also with the inappropriate formulation of the sentences (sometimes it is clumsy and repeatedly said). The authors need to address this and reduce the unnecessary details significantly in order to maximize the scientific impact of this study.

We have attempted to remove all duplicated information. We certainly agree on the removal of figure information from the main body of the paper. As for the level of details in the presentation, this is more subjective. Some readers may want a lot of details on the data processing and analysis. Indeed, both reviews for the paper ask for additional details for specific aspects of our work. Nevertheless we tried to remove the unnecessary ones.

Additionally, the paper should be re-organized a bit for sharing effectively its flow of thoughts with the readers (for e.g. (1) it is far better if authors describes the components of the model (sect. 2.2 + 2.3) after introducing the model (sect. 2.4), (2) I don't follow the reason why authors provide a part of "Discussion" (sect. 2.6) here, before the result sections – Please reorganize it: one option is to shift sect(s). 2.5+ 2.6 to the result sections).

Following the reviewer recommendation, we moved the description of the atmospheric transport model before the description of its input components. In addition, we renamed the "discussion" section to "Analyses and insight for the inverse modeling configuration". Indeed, this purpose of this section is to analyze the results of the direct simulation and measurements comparison to define the inverse modeling setup, which explains why we cannot shift it later.

The second major issue is with the Eiffel tower simulations (EIF) which do not capture the observed variability and gives a model-measurement bias of -13 to -30 ppm that is too large for this kind of applications done in the study. Since 300 m tower measurements is capable of giving valuable information about larger area and is easier to represent in the model in terms of (local) near surface process, it is important to check further why there is such a big bias in the model consistently (or where and why the model fails to reproduce the observations). Knowing that improper vertical mixing produces considerable model-measurement mismatches in mesoscale models, more careful analysis is required in this aspect to establish robustness of the current inversion method (relevant for other sites as well), before utilizing it in the inversion or applying "quick fix". It is not clear to me from the text or figure whether the largest mismatch occurs in the diurnal cycle.

I "assume" that it is mostly during night where the model might already "see" free-troposphere in that height. This could be examined by a comparison of simulated mixing heights with observations (ceilometers- Jussieu?). I do not see such a remarkable improvement in results after applying

“quick fix” (removing EIF observations from the inversion system). This is something to be noticed seriously. Without this kind of analysis and quantification or qualification, I would think that these results are somewhat preliminary to be published.

Note: we have a -13 ppm afternoon bias in November-December at EIF and several -30 ppm hourly mean misfit values in November-December or October-November (figures 8 and A5). However, the afternoon bias in October-November at EIF is -5 ppm so that we do not have a “-30 ppm bias” during the afternoon at EIF, i.e. during the period used for the inversions.

Note: Analyzing in depth the errors at night is actually out of the scope of this study. Like most of inverse modelers, we have focused on the assimilation of afternoon data given the lack of confidence in the transport models during the rest of the day, when the mixing layer is not so well developed. Still, the paper displays some digressions regarding model – data misfits at nighttime in view to confirm that despite the urban heat island effect, the model does not seem to behave better at urban sites than at other sites at night. Figure 2 below (within the answer to the second reviewer) provides some more insights about the diurnal variations of the mole fractions misfits at EIF.

We understand the concerns of the reviewer regarding the misfits at EIF. However: the fact that misfits are very large at EIF is not particularly surprising given the local heterogeneity and amplitude of the CO₂ emissions in cities, and given the complexity of local transport processes associated with the urban canopy. While one can expect that the signature of such processes and emissions have sufficiently been diffused along the path to GIF, GON and MON stations at the outskirts of the city, they can impact a urban site such as EIF even at 300 m agl. Actually, most of the high misfits are obtained when the wind speed is low (Figure 2 in the answer to the second reviewer), or when the vertical mixing is low (in December; see the differences between figures 8 and S5). The simulations conducted in summer demonstrate much reduced model-data misfits (current analysis for a forthcoming paper that uses applies the method to a full year of data). With its 2 km horizontal resolution, its tendencies to numerical diffusion, and its lack of information regarding the 3D structure of the urban canopy or of the injection heights for local CO₂ sources, an Eulerian model such as the one used here cannot be expected to model urban CO₂ as well as rural CO₂.

The use of urban sites such as EIF for atmospheric inversion will likely require long term research for the inverse modeling community. Previous results obtained at MeteoFrance by Lac et al. (2013, ACP) using a high (2 km) resolution meteorological model that includes urban parameterizations, and validated against local meteorological measurements, also show high model-data misfits at EIF, similar to those found in the present paper. The same study finds smaller misfits at GON, MON and GIF like in this study. Mc Kain et al (2012, PNAS) also indicated that their transport model had a poor skill for modeling urban sites so that they preferred to ignore the information content of the measurements regarding the absolute emissions, focusing on long term relative changes in the emissions only. These studies confirm that our problems for modeling urban CO₂ at EIF using a 2km resolution transport model were not really surprising and that the urban misfits obtained in this study are presently typical of this generation of models.

Therefore, acknowledging our lack of understanding of the urban CO₂ measurements, we prefer to limit the scope of this study to (1) the assimilation of data at the semi-rural sites (which is closer to a more traditional inversion framework than the assimilation of urban measurements) (2) mentioning issues at urban sites to stress a new matter of concern and study for the inverse modeling community. Indeed, the removal of EIF data from the assimilated dataset is not a “quick fix” in order to improve some statistics of the results. This explains why the reviewer cannot see any improvement in the results when removing EIF (the inverted fluxes could look reasonable even when assimilating EIF). The removal of EIF is rather a precautionary measure taken due to the analysis of the prior misfits, since, as said by the reviewer, they seem too large for the inversion, and since they reveal our need for longer-term studies in order to assimilate urban CO₂ data.

We shall try to discuss these points better in the revised version manuscript.

My last concern is about the concentration gradient method used in the study (sect(s). 2.6+ 5). The discussion on the contribution from the background is somewhat unclear and hard to follow. My first question is how valid these kind of assumptions (on setting measurement sites as background reference stations and rely on simulated wind speed) inside Paris area. Many are different here: orography effect, measurement level etc. Second question is how it is made sure that the background “reference site” concentrations had already “fully” reached in respective “measurement” sites and no parts are missing due to couple of other reasons, which depends on wind speed (advection) in addition to wind direction. In my understanding, this gradient method performs some filtering of the data which is difficult to model (in terms of both transport and flux); however there is high probability that these “difficult to model” data contains valuable information about fluxes, than just “background noise” as assumed here. As already shown, the negative values in Fig. 9 clearly indicate the failure of these assumptions. One must be more careful in performing these kinds of “filtering”. More valid analysis is required in terms of expected variability of the background, before applying this method.

Although we understand the concerns of the reviewer, we stress that, if the modeling of the transport between upwind and downwind sites was wrong enough to invalidate the “gradient” approach, it would also invalidate the attempt at inverting the fluxes assimilating concentration measurements individually:

- The main objective of the gradient method is to focus the inversion on city scale emissions by removing most of the upwind signal from the measured and modeled concentrations. The upwind signal is made of remote fluxes both from the boundary conditions and from fluxes within the domain but outside the city that are poorly represented by our model input data and that cannot be well constrained using the network of available measurement stations. Due to technical limitations, present inversion systems cannot easily account for the large scale errors arising from uncertainties in the remote fluxes. Consequently, they tend to transfer some of these errors into the inverted city emissions. By trying to remove the signature of the remote fluxes in the “observation vector” of the inversion system, the gradient method attempts at overcoming this problem. Our approach relies on the assumption that, due to atmospheric diffusion, the signature of remote fluxes upwind the city is sufficiently homogeneous in space (horizontally and vertically) and time over the path through the city from upwind to downwind sites (both located within the afternoon PBL), so that the main part of such a large-scale signal can be removed by the gradient method. The term “background” which improperly corresponded to this “large scale upwind condition” was clearly misleading and we have removed all the occurrences of this term.
- Ideally, we would have defined and selected the gradients so that they would have been fully consistent with the measurement of the differences of concentration for air parcels before and after their path through the urban area (the model accounting for the advection, dispersion and mixing with other air masses along this path). By doing so, we would have constrained the city emission budget more directly. However, this would have required a complex gradient selection based on a Lagrangian monitoring of the transport and dispersion of air masses from upwind sites and few cases may have been selected. We rather select gradients based on the selection of a relatively wide range of wind directions at GIF or GON/MON centered on the GIF – GON/MON direction, assuming that for all these wind directions, the upwind and downwind sites bear the same signature of the remote fluxes and that their difference is well representative of the city emissions. We also ignore the time lag needed to transport information from upwind to downwind sites through the city by computing spatial gradients between concentrations at a given time. However, given the wind shear in the PBL, the exercise of deriving such a time lag as a function of the wind speed would be vain. Typical wind speed over Paris at 700 m agl is 7 ms^{-1} (i.e. 25 km h^{-1}) and the distance between GIF and GON is approximately 40 km so that air masses take, on

average, less than 2 hours to travel between the 2 sites at this height. Given that we select gradients for large wind speeds and that we correct for 6-hour mean fluxes, we assume that the assimilation of gradients of concentrations measured during the same 1-hour window is a minor issue.

- Several results support the above-mentioned assumptions even though the reviewer is right about the existence of counter-examples (negative downwind-upwind differences) due to the use of a large range of wind directions used to define the “North-East South-West” direction. Indeed, the model predicts that the signature of boundary conditions vanishes in the gradients that we defined (see the strong decrease of the BC+EDGAR signal between figures 8 and 10 –top row- as an illustration of this decrease) and the scatter plots of modeled-measurement mole fractions clearly indicate that the gradients are better modeled than the individual measurements (compare Fig 8 and Fig 10).
- If the modeled wind directions, speed, boundary layer height... were wrong, this would impact the assimilation of individual concentration measurements as well as the assimilation of gradients. There is no reason to believe that this impact would be larger when using the gradient method. This approach does not exacerbate the need to rely on the meteorological forcing.
- Regarding the loss of potentially valuable information when selecting the data: if we had a method with some potential for filtering this valuable information we would assimilate the corresponding data. Unfortunately, at this stage, we do not rely on the inversion framework for separating the signal from the city and the signal from remote fluxes when the wind does not blow in the (approximately) SW-NE direction. Given the low confidence in our estimates of the remote fluxes, the injection of large errors in the inversion due to the assimilation of such data is thus far more likely than the ability to get an added value from the assimilation of all the data. We should not be too optimistic about the ability of the inversion to separate the influence of remote and target fluxes based on the cross checking of all the sources of information (measurements, prior fluxes, variations as a function of the wind...). This ability strongly relies on the set-up of the inversion parameters by the inverse modeler himself.
- Finally, we only control the total CO₂ emissions of the city, not its spatial distribution. Our system could be driven wrong if assimilating information about fluxes over a small part of the urban area which would not be representative of the full city (i.e. we could generate aggregation errors). The selection of gradients when the wind speed is significant and in the SW-NE direction is a way to ensure that the signal at the stations bears the signature of fluxes throughout the city and that use measurements in the core of the city plume. It should minimize the impact of aggregation errors.

We have tried to clarify these points in the paper. To conclude, we do not claim that this “gradient” approach is perfect. We claim that it yields some improvements compared to the assimilation of all data, at least for our specific configuration, and that it is a valuable step forward in the adaptation of traditional inversion methods (used for the monitoring of natural fluxes at large scale) for their application to the monitoring of city emissions.

I recommend authors to address these issues (+ specific comments below) before publishing it in ACP.

Specific Comments:

Fig. 3 + 4: Indicate spatially averaged values?

The legend of Figure 4 explicitly indicates that the values are for the full Ile-de-France area. We have added this information in the legend of Figure 3.

Fig. 5: a bit disappointed to see that y-axis is not properly set to include the “peaks”. Please redo the figure with proper y-axis setting.

Very large concentration peaks are observed during the night. As explained in the text, the flux inversion does not use nighttime values. Thus, we choose a scaling that preserves a good dynamic for the afternoon concentration, although we feel useful to also show the large variations observed during other time periods.

Fig. 6: “Note the weekly cycle with lower values during Saturdays and Sundays” It is not obvious in the figure. Please mark these days in the time series.

The emission weekly cycle is quite clear in Figure 6 (blue line), with 5 days of high emissions and two days with lower values. We have nevertheless made Sundays explicit on the figure for easier reading of the weekly cycle.

p. 9653: AirParif inventory: Did this study use hourly emissions? Fig.3 shows only week days, Saturdays and Sundays. Please make it clear.

As explicitly said in the legend of Figure 3, it shows the hourly CO₂ emissions for the typical daily cycles of a weekday, Saturday and Sunday. The Airparif inventory that we use considers that all weekdays (Monday to Friday) show similar emissions. The first sentence of 2.2 says that the AirParif inventory is hourly. Also in 3.6: “Note that the AirParif inventory has a 1 hour temporal resolution”.

p. 9654: “Figure 2 shows an example” October or November? text and Figure details differ in terms of period. Please check the whole manuscript.

Corrected

p. 9655: “The direct simulations and the atmospheric inversion thus account explicitly for the biogenic fluxes” Not clear. Please rephrase.

The sentence is not necessary and we removed it.

p.9666: Sect. 3.6: Not clear how you determined the matrix H.

The two sentences have been modified in the text to make it clearer: “We use the atmospheric transport model to compute the impact to the mole fraction of each surface flux (156 in total) corresponding to an element of the control vector. The 4D mole fraction fields from each of these simulations are then sampled at the place and time of the atmospheric observations. This provides the elements of each column of the H matrix.”

p.9667: “One may then blame” “blame” -> Use better word

Sentence changed to “The boundary conditions are then most likely responsible for the misfit during that particular event.”

p.9668: “One notes that the posterior estimate of the afternoon NEE” Do you have an explanation why it became slightly positive?

This is the result of the inversion system which led to this solution. However, the error bars are still large (four times larger than the posterior flux estimate) and do encompass negative values. The inversion system does not rule out a negative (uptake) flux during the afternoon, but, on the opposite, rather leaves a full uncertainty on the sign of the flux. The correction from a negative value in the prior to this nearly null value in the posterior can make sense in late fall-winter, and for the corresponding 6-hour time window.

p.9670: “The second line shows” “line” -> panel

We have made a correction but use “row” rather than “panel” as it orders from top to bottom rather than from left to right and the comment applies to many panels (all in second rows)

p.9673: “we have set a lower uncertainty for the gradient” “lower uncertainty” -> how much (%)
The sentence quoted by the reviewer explicitly refers to section 3.5 where this is discussed.

Anonymous Referee #2

Overview: Beron et al.’s manuscript is focused on attempting to use atmospheric observations of CO₂ in the Paris region to constraint CO₂ emissions. They present measurements collected at 5 sites in and around Paris, and perform two different flux inversions; one using all five sites and inverting for fluxes, the second approach inverting the gradient between sites focused on three locations. ECMWF winds are used to drive the CHIMERE model, and a linear Bayesian approach is taken to optimize prior inventoried emissions. The authors report optimized fluxes for two 30-day periods, with key findings that the Eiffel tower station to be poorly represented by their framework and the gradient flux method appears greatly superior to inverting absolute concentrations.

We fully agree with this summary, and thank the reviewer for his/her work and suggestions

Overall this paper is appropriately placed in ACP, is interesting, and much of the analysis and conclusions are sound. As currently presented, the manuscript reads more like a class project than a complete scientific analysis, and improvements on specific aspects of the analysis in addition to organizational and text refinements are needed, after which publication in ACP would be appropriate.

We have tried to improve the manuscript following the reviewer suggestion. The structure of the manuscript has been slightly changed as described in our answer to reviewer 1.

Major Issues: The two larger issues I have are:

Presentation/style. This is perhaps a more minor comment, but would require some reworking of the manuscript. The manuscript reads more like a report than a completed analysis/paper. There is some meandering in the language leading to some repetitive sections and causing some confusion with the reader. A more focused rewrite emphasizing on the specific important elements would be preferred.

We have tried to simplify and strengthen some presentations and remove some duplications. However, for the overall presentation of the work, we have attempted to provide a clear explanation of the inversion methodology. This description, we feel, is necessary for the readers that are not familiar with the flux inversion and, given our topic, we target a much wider audience. We hypothesize that the reviewer is familiar with the inversion framework, so that he felt that some description were unnecessary.

2) The Eiffel tower data-model mismatch. One would think this would in fact be the easiest data to simulate, and also would be the most representative and useful in the inversion. The finding that this site seems to be inconsistent with the model is worrisome and also would imply that the inversion system is not properly representing vertical exchange, therefore biasing the analysis with the surface sites. This disagreement needs to be discussed in more detail and better understood. Is the disagreement greatest at night/during the day? Does it appear to be link to erroneous mixing heights? Is there evidence of persistent eddies developed around the tower causing elevated signals not represented in the model? There is currently very little discussion on this, and we need to see more discussion and data to better understand the potential reasons for failure at this site and possible implications for the inversion system.

This likely should be a little section to itself including figures. Depending on why the mismatch occurs, it would impact the conclusions made about the remainder of the inversion system.

We agree with the reviewer that the measurement-model mismatch at the Eiffel tower is worrisome. Indeed, it is commonly admitted that high towers provide more valuable data for flux inversion that

near-surface sampling. However, the Eiffel tower is specific in the sense that it is located within a large urban area i.e. above a very strong and heterogeneous source and subject to local atmospheric transport processes that are complex due to the urban canopy. Please, see the detailed answer to the second general comment by the first reviewer that is similar to this one.

This answer recalls that

- there are reasons to assume that local sources and transport can raise large misfits at EIF without impacting the skill for modeling sites at the edge of the city due to atmospheric diffusion
- some of the results (the larger misfits at EIF in December and when the wind speed is lower) and previous studies (Lac et al. 2013, ACP; Mc Kain et al. 2012, PNAS) support this hypothesis
- in answer to some of the second reviewer's minor comments we display some analysis of the CO₂ or wind model-data misfits at EIF and Saclay (near GIF) respectively (Figures 1 and 2 below). These diagnostics highlight the challenge for modeling the urban CO₂.
- our study is careful about acknowledging our lack of understanding of some of the CO₂ measurements. Removing the EIF data from the inversion is thus a sanity measure. The assimilation of urban CO₂ measurements will likely require long-term studies.

The new manuscript better clarifies these points.

2a) The concentration gradient method. It is interesting that this seems to work better, and the message of importance of constraints on incoming CO₂ levels is very important many networks have been designed more recently (see LA) with boundary values kept in mind. However, the gradient method applied in the inversion here is only one method to constrain this, and a very simplistic one at that. It is important to make it clear that there may be other methods to constrain the incoming CO₂ concentration that are not quite as limited in their application.

We certainly did not imply that the “gradient approach” we have presented in the manuscript is a universal technique, or the best one. As stated in the answer to the third general comment from the first reviewer, we simply think that this is a useful step forward, at least for configurations that are similar to ours, for the adaptation of traditional inversion method (used for the monitoring of natural fluxes at large scale) to the monitoring of city emissions.

This method may not be relevant for different inversion configurations with different city surrounding and emission spread, and local meteorology. In particular, this is better adapted to cities where the emissions are concentrated over a narrow area and transported through a plume outside the city without any local accumulation due to strong topography or typically low wind speeds. We agree that in LA, where CO₂ accumulates over the city due to the surrounding mountains, one could hardly rely on a downwind-upwind gradient method.

Please, see also our discussion on this method in answer to the 3rd general comment by the first reviewer. We hope that this demonstrates that even though it sounds “simple”, this method is not “simplistic” in the sense that it does not rely on stronger assumptions than the assimilation of individual measurements, and in the sense that it definitely strengthens the inversion framework.

We have tried to highlight it better in the new manuscript.

Minor Issues:

Pg. 9649 line 10. Would be appropriate here to add citation for the study preceding the McKain paper (Strong et al., 2011, JGR) and for satellite attempts at the problem (Kort et al., 2012, GRL).

We agree. These references have been added to the introduction

Pg 9649 line 27: Would be appropriate here to cite network footprint/ design studies calculated for Los Angeles, both of current and future observations (Newman et al., 2013 ACPD; Kort et al., 2013, JGR).

We disagree. Although these papers are of value, they are not directly connected to the discussion here. They have not been included to the manuscript

Pg. 9651 lines 1-2: It is a simplification that is not necessarily accurate to state the atmospheric transport modeling will be simpler for Paris the flat topography and winds lead to more of a plume like structure, which may in some senses be simpler to simulate, it necessitates a much tighter requirement on the simulation to get the plume location and dilution very accurate relative to the observations, which may be more different to simulate than a city that accumulates more of a dome due to surrounding topography.

On the first order, it may be expected that the limitation in spatial resolution of the atmospheric transport models is less an issue over flat terrain than where the topography is complex, at least regarding the representation of measurement sites, but also regarding the modeling of transport from the flux to the sites. The plume structure also bears the advantage that the model relies less on the history of the emissions and of the transport to model the concentrations than for cases where CO₂ accumulates.

Furthermore, there is no “dome” of CO₂ over a city strictly speaking in the sense that lateral “leakages” necessarily occur e.g. through the mountains surrounding the city or at their edge. We can debate whether this can significantly alter the inversion. But we believe that the first points raised above give a major asset to flat topography and they are clarified in the text.

Page 9653 line 3: is this miswritten? Or is the repeatability not know “expected to be better than 0.3 ppm”

The sentence has been corrected to “They have been regularly calibrated against the WMO mole fraction scale (Zhao and Tans, 2006) so that measurement accuracy to the WMO-X2007 scale is estimated to be better than 0.38 ppm. The instrumental reproducibility is better than 0.17 ppm on the 5 minute average measurements available from the CO₂-Megaparis stations, but the temporal averaging to the hourly-mean values used in this paper leads to precision much better than the accuracy.”

Pg. 9653 line 10: Are these local or UTC times? Is there daylight savings?

These are UTC times which, as indicated in the next sentence, are very similar to solar times in Paris. Daylight savings apply to legal time that do not apply here. The emissions (AirParif inventory) are indeed provided for legal time, and the data processing does account for the difference between legal time and UTC time, including daylight saving. The following sentence was added at the end of section 2.2 : “The AirParif inventory is provided as a function of legal time, and we have accounted for the time shift between legal time and UTC time, including the impact of daylight saving.”

Pg. 9654 Line 4-5: I am surprised by the day choices. In the US we know Monday and Friday are both distinctly different from mid-week or weekend, and I would expect the same in Europe. I would think the subsection of simple weekday would lead to some bias errors those days.

There is no significant difference in France between weekdays, and AirParif only considers weekdays, Saturdays and Sunday in its inventory, as described in the manuscript.

Section 2.3: Is there any validation at all for the biogenic component? Can you leverage anything from the Lac et al paper? Are these urban NEE or rural? Is there any city distinction? We need more description of what is done here, since the simulated NEE is actually important in the inversion result.

The C-TESSSEL simulations are “state of the art” but we have not attempted any validation of their values. As described in the manuscript, the C-TESSSEL simulations have a 15 km resolution and uses a land use map at even higher resolution through a tiling approach. They have been interpolated to the CHIMERE grid. Figure 4 shows the aggregated fluxes over the Ile-de-France

region but we use NEE estimates over the whole CHIMERE domain. Although we agree that the biogenic flux is a significant contributor to the CO₂ signal, it is not clear what can be added in our manuscript. The details of the C-TESSSEL modeling are provided in Boussetta et al. Note also that the gradient method reduces the sensitivity of the results to the uncertainty in the NEE.

Pg 9656 Line 8: The lack of an urban scheme is potentially worrisome. We know from observations that Paris has an elevated pbl relative to its surroundings. Is that represented in the current simulation? Is there a justification for no urban scheme inclusion?

We explicitly recognize this limitation of our study in the paper. We have no urban scheme in the atmospheric transport model that is used for this paper. On the other hand, we have compared the Chimere simulations driven by the meteorology generated by Lac et al, 2013, ACP (that uses Urban parameterizations) to the Chimere-ECMWF simulations and found that the latter leads to a better agreement with the CO₂ mole fraction measurements even at EIF (results not shown in this study). The urban parameterization is not the only difference between the two simulations so that one cannot conclude about the cause of the improvement. Nevertheless, the urban scheme does not appear to be a key driver of the model performance for CO₂.

Pg. 9658 Line 20-23: This is very important boundary conditions matter a lot. This in fact is part of the message of this paper.

We agree. This message is in the abstract, which makes it a key point of the manuscript

Pg. 9658 Line 25: This comment doesn't make sense to me. It should matter what the wind direction is there are very different background concentrations/far upwind sources in different directions of Paris.

We agree. The comment was wrong and has been removed. Indeed, we select the wind directions in the gradient approach. Of note is also the removal of the term “background” from the manuscript since it potentially raised such a confusion.

Pg. 9659 Line 1: which is bigger by 30 ppm? It is very hard to discern in the current figure.

With the pdf file, it is easy to zoom. The diamond symbols are very easy to see. We feel the reader can verify, for instance in Figure 5, that the difference between the model (green) and the measurement (red) are very large for the Eiffel site on Dec 3rd and Dec 12th as indicated in the text.

Line 2-4: We need far more quantitative discussion of this, as opposed to just a couple guesses. We need more data presented and understanding for why this discrepancy occurs.

We agree that the measurement-modeling mismatch at EIF is worrisome. We did put significant effort to obtain a better understanding of the measurement-model discrepancies at EIF. However, we have not been able to derive a satisfactory demonstration from this effort. A full analysis of the signal at EIF and the discussion of the impact of modeling errors would be a paper by itself. We feel it is out of the scope of the present paper that attempts an estimate of the Paris emission using the tools that are readily available. See also the answer to the general comment regarding this and Figure 2 below.

Line 12: specify for this flux inversion not necessarily true of all inversion systems

Agreed, although it is a very common feature of atmospheric inversion attempts as most/all models have difficulties in representing atmospheric concentrations in stable conditions. We have changed “are not appropriate for the flux inversion” to “are not appropriate for our flux inversion”

Line 14: this has not really justified this decision. It is more of a qualitative support.

We explain why we made that decision. Indeed, the decision was made on qualitative rather than quantitative arguments.

Pg. 9660 Line 10+: This would impact other sites as well?

Eiffel is very specific in the sense that it lies in the heart of the city where emissions are very strong and heterogeneous. The other sites are more rural, further from the emissions. A direct consequence is that the atmosphere has sufficient time to homogenize within the boundary layer when it reaches the station. This certainly explains why the measurement-model agreement is much better for the monitoring sites other than Eiffel. Although we agree this is qualitative, there are strong indications that the modeling framework can reproduce the rural and semi-rural mole fractions even if it does not reproduce properly those at Eiffel.

Pg. 9662 Line 10-13: Why this selection for the NEE? One might expect NEE to change fairly drastically day to day given variation in say PAR.

NEE can change drastically from day to day but the C-TESSSEL estimates does model day to day variations. Inversions of the NEE usually assume rather large autocorrelation timescales for the uncertainty in the NEE, typically 1 month (e.g. see Broquet et al. 2013, ACP). Such timescales have been derived from analysis of the misfits between the typical prior estimates used for the NEE (from vegetation models) and eddy covariance measurements of the fluxes (such as in Chevallier et al. 2006, GRL and 2012, GBC). For one month experiments such as here, defining a 1-month correlation length scale for the prior uncertainty in NEE is roughly equivalent to rescaling the NEE over this month without adjusting the day to day variability.

We agree that the errors in NEE have also components that vary day to day. The analysis of Chevallier et al. 2012 indicate that they should be less significant than the 1-month scale error and inversion system generally account for one temporal scale of the uncertainty only. Still, we agree that such day to day errors can impact the inversion of the city emissions. The gradient approach in the inversion partly cancels out the impact of this error.

Line 21: It would be nice to directly use observations to constrain this.

We agree, but there is an insufficient number of observations around Paris to constrain fossil fuel, biogenic fluxes and boundary conditions independently. A selection of independent variables must be made. We have developed a system with some choices that appear reasonable to us, although they may seem arbitrary to others.

Pg. 9663 Line 19-27: Worth noting in this portion of the text that the LA network and the updated INFLUX network are designed specifically to account for the upwind boundary conditions.

We have not added this information here, but references to INFLUX and the LA network are provided in the introduction

Pg. 9664 Line 11-22: I understand there are some arbitrary choices made here but I would like to see somewhere (perhaps a supplement) that explores the impact of these choices how a range of selections would impact your findings

We have done this. Fig S-3 in the supplementary is the result of one of these analyses. All the main findings of the paper (i.e. those that make it to the abstract) are not changed when changing the correlations within a reasonable range of values.

Pg. 9665 Line 10: "We deduce" How? Where? You need to tell & show us.

We recognize that the sentence was insufficiently detailed and it has been rewritten. When we assigned the observation errors, we ensure that the projection of the prior uncertainty of the fluxes in the concentration space plus the observation error do fit the actual prior misfits between the model and the data (i.e. that the term $\text{HBH}^T + R$ is consistent with the RMS value of the prior model-data misfits; see, e.g., Desroziers et al. 2005, QJRMS). We also verified that the mean of the product between posterior and prior model-data misfits are consistent with the observation errors, using the diagnostic of Desroziers et al 2005, QJRMS. These two diagnostics should make the assigned prior and observation uncertainties realistic.

Line5-15: I am a bit worried on this approach as by definition it down weights where observations and model simulations disagree which may indeed be due to problems with Emissions exactly what you are most interested in!

The diagnostic accounts for the projection of the uncertainty in the prior fluxes into the concentration space to avoid such a problem. This is not an “approach” based on our own assumptions regarding the observation error. The calculation strictly relies on the hypothesis made by the inversion system that observation and prior errors are independent and that observation error encompasses all the sources of model data misfits that are not accounted for in the prior error (see also Desroziers et al. 2005). If we underestimate the prior uncertainty, then this diagnostic may overestimate observation errors and we may lose some useful information, but on the opposite, if we overestimate the prior uncertainty, this diagnostic will underestimate the observation error and the inversion will revert some of the model transport errors into the flux corrections. This is a traditional issue for inversion systems. Being aware that the assumptions underlying the inversion framework are not perfect, we think that it is better, for sanity measure, to use an R that is conservatively too large rather than a R that is too small.

Pg. 9666 Line 10: Is there any error analysis of the winds? Do we have any information on how accurate winds are? PBL heights? Or any other meteorologically critical variable for these comparisons? This would be important to add and include.

We do not feel that the validation of the meteorology is within the scope of the present paper. We use the state-of-the-art ECMWF meteorological fields at 15 km resolution which have been evaluated elsewhere. Most atmospheric inversion studies do not analyze the quality of the input meteorological data. Nevertheless, we have performed some analysis of the wind field with results that are similar to those presented in Lac et al., 2013, ACP. The text now discusses briefly these results. The two figures below show the statistics of wind speed and directions at the Saclay meteorological tower (very close to GIF). We performed our analysis based on these measurements because (i) the data were readily available to us and (ii) they are not used as input to the ECMWF analysis and can therefore be used as an independent variable.

The wind speed and direction do not show significant biases. The typical wind speed error is 1.5 ms^{-1} and increases only slightly with the actual wind speed. Thus, the relative error on the wind speed (which is to the order linked to the relative error on the retrieved flux) is much larger for low wind speed, which provides further justification for their rejection in our inversion setup. As for the wind direction, the typical error is very large for low wind speed, but is 15° or less for wind speeds larger than 5 ms^{-1} .

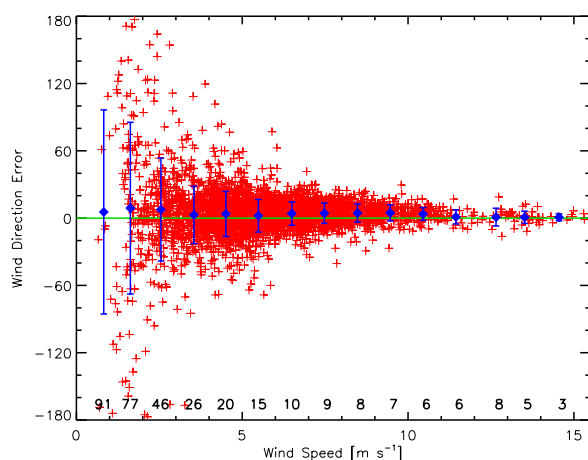
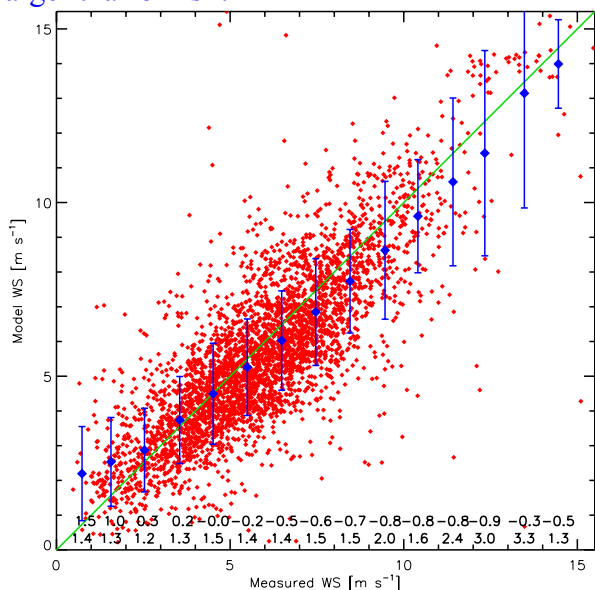


Figure 1: Scatter plots of the measured and modeled wind speed and direction at the Saclay tower (100 m agl). The left image is for the wind speed. For each 1ms^{-1} bin of wind speed, one provides the bias and standard deviation (shown in the image and quantified by the numbers at the bottom of the plot). The right image is the wind direction. Clearly, the wind direction accuracy depends on the wind speed so that the wind direction error is shown as a function of wind speed. The symbols and error bars show the bias and standard deviation for each 1ms^{-1} bin of the wind speed. The numbers at the bottom give the standard deviation.

Pg. 9667 Line 20-21: this is also likely due to poor transport representation of accumulation instead of local sources. Is there an analysis of the footprint of the sites? This would seem a critically useful figure. How do we know the representativeness of the sites? The changing daily footprint could better explain agreement/disagreement. This also might be helpful in understanding the limitations of the EIF site.

We use a threshold of 2m s^{-1} on the wind speed at this stage of the paper. Therefore we doubt that we can really speak about CO₂ accumulation. Many hypotheses can be raised regarding the difficulties in simulating of urban CO₂, and the present analysis and available tools can hardly help discriminating between them.

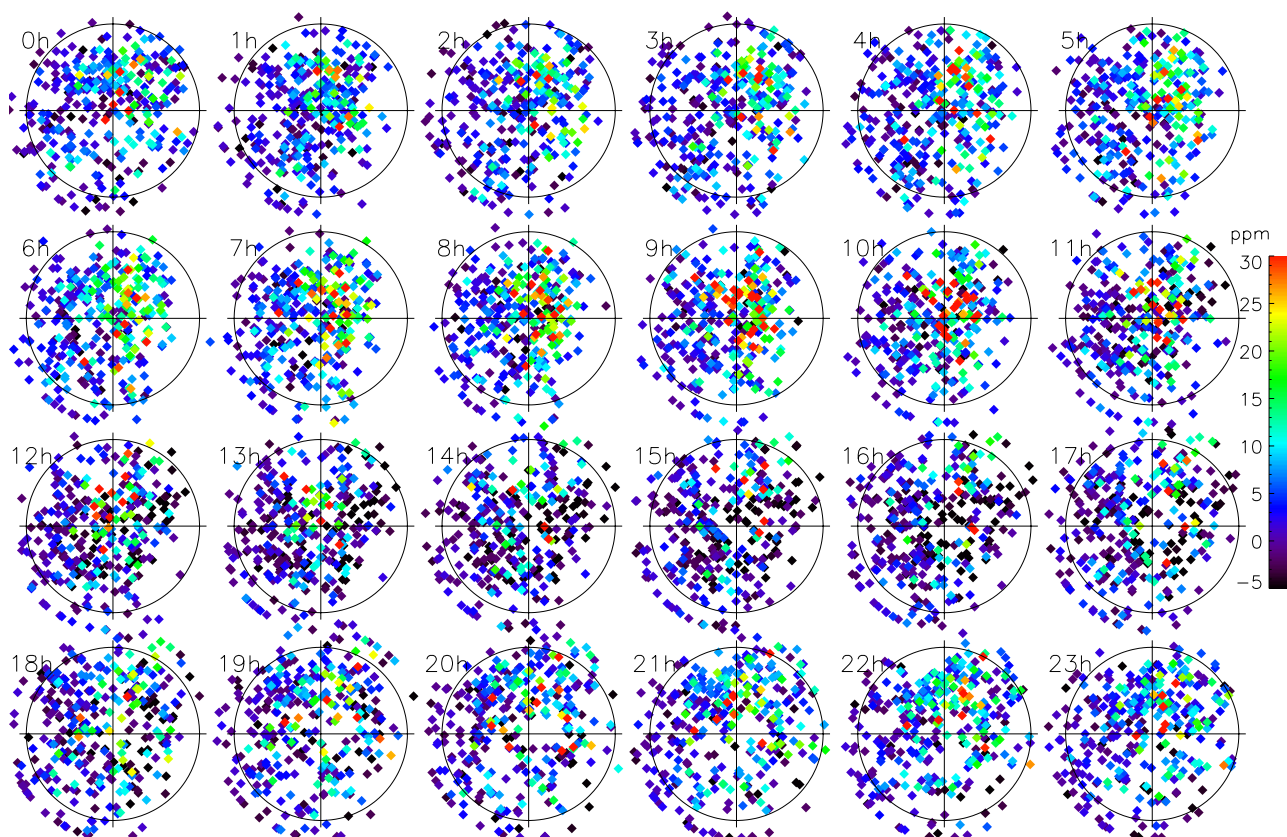
Regarding the footprints: the reviewer likely refers to the modeled sensitivity of instantaneous concentrations to the grid cell / hourly fluxes. It would be used to identify the potential source for a given hourly misfit by looking at grid cell / hourly fluxes to which the concentration would be highly sensitive to.

We could compute such “footprints” with the adjoint of CHIMERE. However, footprints are built with the model, so there is no reason to believe that they will bring critical insights on the model data misfits. If misfits occur because of a wrong wind direction or speed, wrong horizontal diffusion, or from wrong vertical mixing, one cannot interpret them with footprints estimated with these wrong parameters. In that case, the analysis of the misfits as a function of measured wind speed and direction such as in the figure below raises more insights about the sources of errors.

Footprints do not raise any insight about the model representativity errors that could dominate the model-data misfits in a urban environment.

If the misfits arise from subgrid scale sources (not represented in the Airparif inventory) or from a wrong emission distribution of the emissions at 2km resolution in the Airparif inventory, and if such sources of errors have few locations, one may get an idea on these locations by looking at the direction of the advection demonstrated by the footprints when the misfits are high. However, footprints are only estimates of sensitivity and would not indicate the actual location of the error sources along such a direction.

The figure below is an analysis of the measurement-model misfits at EIF as a function of wind and time of day. The figure clearly shows that the best agreements (purple and black dots) are observed during the afternoon. On the contrary, the largest misfits are mostly found for low wind speed and during the morning. Our interpretation is that CO₂ accumulates during the night and reach the instrument at the top of the Eiffel tower when the boundary layer thickens. The misfits tend to be larger when the wind blows from the North-East. This is interpreted as a default of the boundary conditions as the air mass originates from the Benelux region which is a strong emitter of CO₂.



Pg. 9669 Line 9-10: *That is assuming the prior spatial distribution is accurate.*

Agreed. We have added the following sentence to stress this point : “However, this attribution relies on the a-priori spatial and temporal distribution of the fluxes. These distributions are affected by errors that are not explicitly accounted for in the inversion set-up even though our diagnostic of the observation error should encompass it.”

Pg. 9675 Line 14: *Typo : constrain should be constraint*
Corrected, thanks

Pg. 9678 Line 29-30: *Specify this statement is just for Paris.*

The statement is about the need to rely on prior information on the daily cycle of the emission to constrain the daily total, because night time mole fraction measurements are not used. As acknowledged earlier, this does not apply to all inversion frameworks. However, this concerns a majority of the inversion cases. The text has been corrected.

Figure 2: For what time of day is this?

The legend says “for a week day” and the unit is $\text{gCO}_2 \text{ m}^{-2} \text{ day}^{-1}$. We feel it is clear that the figure shows the day-total emission. We nevertheless added “day-total” in the caption

Figure 3: Why is home heat higher on weekdays? This seems totally counter-intuitive.

We agree with the reviewer and have contacted AirParif that developed the inventory used in this study. It appears that the inventory uses the same weekly profile for the residential and tertiary sector, which is not appropriate. Indeed, in the new version that is currently being developed, emissions from the tertiary sector are larger during weekdays than during weekend while the opposite is true for the residential sector. This point raised by the reviewer demonstrates that current inventories are far from perfect, which does leave room for improvement that might be aided by atmospheric inversion such as that developed here.

Figure 6: Please add more labels on the x-axis (here and elsewhere)

We have tried, but it makes the figure more difficult to read. The version we chose provides all the necessary information, in particular since the tick-marks clearly indicate the week periods. Having more labels does not add any useful information. We have added grey areas to indicate Sundays on the figure.

Figure 7: The large NEE adjustment is worrisome is this large predicted sink actually realistic?

The predicted sink is only during the afternoon, when photosynthesis dominates. The day-total NEE is a source. The C-TESSSEL model has been validated at ECMWF. It may not be perfect locally in the Paris area but it is out of the scope of this paper to analyze the NEE in depth. The inversion corrects the prior sink to a neutral flux which could sound more realistic in winter, but with an uncertainty ranges that includes both source and sink. Clearly, our inversion system is designed to monitor the fossil fuel emission, and there is less confidence in the results concerning the NEE.

Figure 9: Please make arrows bigger and clarify this figure that is currently hard to read/interpret. Between what station and what other station? Hard to discern from the caption.

We changed the first sentence of the caption to “Time series of the mole fraction differences between a station (Y-axis label) and another one used as a reference (either GIF or GON) and selected based on the wind direction (see section 3.3). ”

Figure 12: I see the contrast between this and Figure 7 as indicative the statement that NEE and anthropogenic sources can be spatially distinguished to be erroneous.

The contrast highlights that in the case where we do not use the gradient method, the distinction is more difficult (even though we cannot demonstrate it based on these figures since they only show a sensitivity of the separation between anthropogenic and natural fluxes to the change of method). It does not imply that this distinction is poorly made when using the gradient method. And as highlighted in the text, the diagnostic of the ability to separate the 2 types of fluxes is based on the inversion set-up only which does not imply that it applies to the real world.