We would like to thank Referee #2 for the comments concerning our manuscript. Please find below a point-by-point response (in blue) to each of the comments raised by the reviewer (in black). All the figure numbers correspond to the revised manuscript.

Anonymous Referee #2

Received and published: 3 August 2020

General comments:

The authors attempted to determine the uncertainty sources for the modeled CO_2 concentrations over Paris, France, using a set of WRF-Chem simulations varying with physics-based transport, fossil fuel emissions, and CO_2 boundary conditions, for 2016. They mainly focused on the impact of PBL schemes and with the combination of the urban canopy models, two fossil fuel emission inventories with /without hourly variability, and two global models as boundary conditions on the modeled CO_2 in comparison with the ground-based in-situ CO_2 measurements. Their results show that model-data mismatch maximizes in the nighttime so they recommended the readers to discard the model-data misfits and use afternoon measurements for inversion. This is not new, and I believe that is what we do in atmospheric inversion.

Regarding this specific item, the point here is the assessment of such a traditional practice in global to regional scale inversions to the specific, fast-growing and more recent field of inverse modeling i.e. that of urban CO_2 emissions based on ~1km resolution transport models, while the transport conditions and modeling skills over urban area can highly differ from larger scale transport conditions.

They also found the boundary condition could cause large differences at the synoptic scale and suggest exploring more about the influence of the boundary condition on the inversed results and suggest using additional observation to constrain boundary conditions. This is also not new. The authors are aware of these points because they cited those papers. So, I failed to locate the novelty of the work that brought into the community. The authors, in my opinion, have repeated some of the previous studies without extending the science further.

We thank the reviewer for giving us the opportunity to better explain the novelty of our results, and to better position our study with respect to similar one. In fact, few studies performed a detailed analysis of different sources of errors for modeling urban CO_2 similar to the one presented in our study (Table R1). Here we present original work for the city of Paris with a deeper analysis on the concept of assimilating cross-city gradients and evaluates the inversions strengths and weaknesses with the results from a full year worth of CO_2 measurements at 8 in-situ stations combined with the meteorological measurements, a sophisticated high-resolution atmospheric transport model coupled with the diagnostic biosphere VPRM model, and a series of sensitivity tests to the main components of the inverse modeling system.

The use of city downwind-upwind gradients for city-scale inversion has been tested for Paris and promoted by a series of few publications (Bréon et al., 2015, Staufer et al. 2016, Wu et al. 2016) cited in this paper. Although the obtained results demonstrate the effectiveness of the inversion system, there are also several aspects concerning its improvement. For instance, the study in Lac et al. (2013), among others (e.g., Kim et al., 2013), demonstrates the potential improvements in the meteorological and atmospheric CO_2 modeling over the Paris region. Therefore, it suggests investigating in more detail whether the urban effects on atmospheric transport modeling need to be accounted for in the inversion of CO_2 fluxes for Paris. Lian et al. (2018) and Lian et al. (2019) attempted at setting up a high-resolution atmospheric transport modeling framework that is more robust or at least more flexible in terms of parameterization than those used in the previous Paris studies to account for the impacts of the urban effects, the biogenic flux and the model physics, which makes it promising to enlarge the set of data that can be assimilated for the inversions of the Paris CO_2 emissions, and in a more general way, to strengthen the inversions.

Moreover, since the publications by Bréon et al. (2015) and Staufer et al. (2016), the Paris CO_2 network has been expanded and relocated since the year 2014, with several new in-situ CO_2 stations combined with meteorological measurements. The present monitoring network, in particular the two newly built urban sites (JUS and CDS), is expected to provide new insights into the urban CO_2 characteristics and the high-resolution atmospheric CO_2 modeling. A full re-assessment of the modeling skill and of the main sources of misfits between the observations and the model was needed on these new bases. This study actually presents a much more extensive analysis of the source of errors in the simulation over a one-year long period, in particular of the meteo-transport modeling errors, and of the skill for simulating the data from the site in the core of the urban area, than in the previous publications. In addition, Paris is a megacity like Los Angeles but surrounded by much more active vegetation in the growing season. In this study, the biogenic CO_2 fluxes were calculated online in WRF-Chem by the diagnostic biosphere VPRM model at 1-km horizontal resolution. We have demonstrated the impact of the biogenic activity on night-time measurements, which was not done before.

In practice, even though the general conclusion converges towards those raised by the previous publications, the study conducted here provides some new error characterization and a range of new detailed insights on the signature of the different types of sources of errors at city scale during nighttime and daytime for the full year period. Our results not only reveal our greatest efforts and current ability to simulate the atmospheric CO_2 concentration in an urban environment, but also prepare a promising way for a better inversion of CO_2 emissions from Paris. Therefore, we believe that the experience gained on what can be done and not done over Paris, could provide useful insights to other cities.

We have modified the text in the introduction section accordingly:

"Since the year 2014, the Paris CO_2 monitoring network has been relocated and expanded with seven in-situ CO2 stations combined with meteorological measurements. The present network, in particular the two newly built urban sites, is expected to provide new insights into the urban CO_2 characteristics. Lian et al. (2018) and Lian et al. (2019) attempted at setting up a high-resolution atmospheric transport modeling framework that is more robust or at least more flexible in terms of parameterization than those used in the previous Paris studies to account for the impacts of the urban effects, the biogenic flux and the model physics, which makes it promising to enlarge the set of data that can be assimilated for the inversions of the Paris CO_2 emissions, and in a more general way, to strengthen the inversions. Therefore, a full re-assessment of the modeling skills and of the main sources of misfits between the observations and the model is needed on these new bases. More specifically, we analyze in detail the model-measurement mismatches so as to identify critical sources of errors that would compromise a high-resolution atmospheric inversion of urban CO_2 emissions in the Paris area. A set of forward simulations of atmospheric CO_2 concentration are performed at 1-km horizontal resolution using the WRF-Chem model (Grell et al., 2005) with different anthropogenic emission inventories, physical parameterizations and CO_2 boundary conditions over the Paris for the 1-year period spanning December 2015 to November 2016. The main objectives of this paper are to provide a rigorous and detailed error characterization of our atmospheric modeling system and to determine the data selection method (i.e. filtering of short-term model errors and local contamination) and CO_2 boundary condition specifications at city scale during both daytime and nighttime over the full year period. We also address the question to what extent these model-measurement mismatches might be reduced and how our proposed diagnostics could be used to provide additional constraints for the inversion of CO_2 emissions at the city scale."

Table R1. Few published studies with the objective to investigate the sources of error in atmospheric CO_2 modeling for city (comparison with this study).

References	City	Objective	Study Period	Measurement	Note
Feng et al. 2016	Los Angeles	Model-data comparison and network design evaluation	One month (mid-May to mid-June 2010)	CO_2 (2 in situ stations), PBL height, meteorological fields	Sensitivity test of physical scheme and spatial resolution
Martin et al. 2019	Washington DC/Baltimore	Analysis of errors in transport and fossil fuel emission	One month (February 2016)	CO_2 (3 in situ stations + 1 rural station), PBL height, meteorological fields	Sensitivity test of four fossil fuel inventory
This study	Paris	Analysis of errors in fossil fuel emission, biogenic flux, atmospheric transport and CO_2 boundary condition	One year (December 2015 to November 2016)	CO_2 (6 in situ stations + 2 rural station), PBL height, meteorological fields	Sensitivity test of physical scheme, fossil fuel inventory and CO_2 boundary condition

Besides, the authors cited a few pilot CO_2 urban studies, such as INFLUX and LA megacity. Both Lauvaux (2016) and Feng (2016) pointed out the significant improvement of using high resolution fossil fuel inventories in simulating CO_2 in the urban environment. Although the authors included two different fossil fuel inventories in the simulation with the variation of the temporal components, I have a hard time following the goal of the experimental design. I thought they would explore the sensitivity of modeled CO_2 to the temporal resolutions when I was reading the methods and section 3.2.1, but the related findings were not emphasized in the conclusion. Why?

This suggestion is well taken. We have added the following sentence in the method section to make the objective clearer:

"In order to investigate the impact of the spatio-temporal distribution (especially the prescribed diurnal profile) of emissions on the modeled CO_2 concentrations, we made a one-month simulation using these two anthropogenic inventories together with their respective temporal profiles (Table 1b). Within the same group of simulations, two more sensitivity tests of the diurnal profile were also carried out by using……"

We also add the related findings in the conclusion:

"Our results indicate that the temporal profile of the heating sector used by the AirParif inventory tends to bear a large uncertainty. It is one of the two major causes that led to the large model-data misfits during the nighttime. In the IdF region, CO_2 emissions from the heating sector are linked to the burning of gas and oil, and electricity consumption. We could expect that a more constant diurnal profile should probably be a better approximation to the truth than the current one. This hypothesis has been further justified by an independent analysis of daily gas use and hourly electric consumption data within the IdF region (unpublished analysis led by a co-author of this study, François Marie Bréon)."

One of the major concerns in the urban CO_2 studies falls in the impact of the biosphere around or within the cities. The results of this study also showed that the impact of the biogenic fluxes is significant and not negligible, meaning that the biosphere is another uncertainty source over Paris. The author can refer to Feng (2019a; 2019b) to construct a set of biospheric fluxes and investigate the uncertainty of the biospheric in the modeled CO_2 as well. Additionally, the authors relied on the VPRM module in WRF-Chem to provide the biogenic fluxes. It's not clear to me that if VPRM has been tuned with flux towers or not. The VPRM parameters in WRF-Chem are fixed and needed to be tuned with the flux towers to have a relatively accurate biospheric flux estimation (Hilton et al., 2013: Hilton et al., 2014). If the authors used the default values for the VPRM parameters in this study, the authors will have to consider the errors caused by the biosphere during the interpretation of the results which is almost impossible to isolate from transport, emission, or boundary condition. However, because of the simplicity of VPRM, the authors can build a set of parameter-based perturbations of the biospheric fluxes via VPRM to address my first concern.

In this study, the VPRM parameters (α, β, λ, and PAR0) have been calibrated using the eddy covariance flux measurements for different vegetation types in Europe made during the Integrated Project CarboEurope-IP (http://www.carboeurope.org/).

More detailed descriptions in terms of the model setup used in this study, e.g. the domain setting, the nudging option and the VPRM model, can be found in Lian et al. (2019) paper. We have referred to this information at the beginning of section 2.1 in this manuscript:

"Details regarding the model setup and the reference data used in the simulations are outlined briefly below and described in Lian et al. (2019)."

For better clarity, we have also added the following statement in the revised manuscript:

"The values of the four parameters (α, β, λ, and PAR0) for each vegetation type used by VPRM have been optimized against eddy covariance flux measurements over Europe collected during the Integrated Project CarboEurope-IP (http://www.carboeurope.org/)."

Given the optimal VPRM parameters have already been implemented in the simulation, we thus feel that there is no need to make a set of user-defined parameter perturbations as suggested by the reviewer.

The authors chose five combinations of the PBL schemes and urban canopy models to study the impact of the transport and concluded that it's difficult to have "good" transport. First of all, what are the rationales the authors believe these two schemes are the key players of the CO_2 urban modeling? Díaz-Isaac (2018) using an ensemble approach pointed out PBL indeed is a major player but ranked No. 2. The most dominant parameterization is the land surface model used in the study. I am aware that the response of the modeled CO_2 to the model physics may vary when the location changes. Have the authors explored the impact of other model parameterizations on the simulated CO_2? This may be also why the model results have such a large bias in this study.

The experiment design and the selection of physical scheme in this study are mainly based on the results of our previous study (Lian et al., 2018). Lian et al. (2018) investigated for the city of Paris, whether the high-resolution WRF model with its various configurations, can provide a good representation of meteorological fields in support of tracer atmospheric transport modeling. A series of numerical experiments (32 simulations) were carried out with the goal of detecting the sensitivity of WRF to its various physics schemes (including 6 microphysics schemes, 3 radiation schemes, 5 cumulus convection schemes, 4 PBL & surface layer schemes, and urban canopy scheme) and nudging strategies (spectral nudging, grid nudging and the objective analysis program OBSGRID). The meteorology provided by WRF was evaluated against both ground-based and radiosonde vertical observations, with a focus on three atmospheric variables (air temperature, wind and PBL height) that are relevant to the CO_2 transport in an urban environment. Our sensitivity tests with different WRF physics schemes show that the wind speed and the PBL height are much more strongly influenced by PBL schemes with respect to other physics schemes. The WRF model together with its urban canopy scheme makes it possible to represent the urban heat island effects. Results in Lian et al. (2018) show that the meteorological variables are generally well reproduced by our WRF setup with the objective analysis and multi-nudging options that have also been used in this study. Meanwhile, it also provides an objective method for us to select the appropriate model physical schemes. Thus in this study, we only carried out the sensitivity simulations with different PBL and urban canopy schemes as they are sufficient to address the paper main question regarding the ability of a configuration of the WRF-Chem model to simulate the atmospheric CO_2 transport over Paris, but also to provide an estimate of the atmospheric transport uncertainty. All the other physics options remained the same in the experiments.

In the manuscript, we already mentioned that:

"These options correspond to those selected by Lian et al. (2018) which showed good performances for simulating near-surface winds and temperatures over the Paris region."

For better clarity, we have also added the following sentence in the revised manuscript to account for the reviewer's comment:

"These two physics schemes were selected as they have a more significant impact on the simulated meteorological variables than the other schemes based on our previous sensitivity study (Lian et al., 2018, Lian et al., 2019), and thus the differences between simulations with these two physical options could provide an estimate of the atmospheric transport uncertainty over the Paris region."

Secondly, the model-data mismatches are extremely large and out of my expiation. For example, the whole year averaged diurnal mismatch can be as large as -10 to 5 ppm at the two urban sites in Figure 5. I found a similar figure in Figure 9 of Feng (2016), even though it's a month averaged value, in which the diurnal cycle from the high-resolution simulation looks almost identical to the observation. What causes the large bias in this work? I would check if any errors caused by other model physics.

Note: Figure 5 is now ranked as Figure 6 in the revised manuscript.

At first glance, the large model-data mismatches in Figure 6 shown in section "3.1 Overall model performance" might be a surprise since the model underestimates CO_2 with a bias ranging from 0 to 12 ppm across stations during the night until around 05 UTC. In fact, a fairly detailed explanation of the two causes of this model-data discrepancy was already provided and justified in section 3.2 of the manuscript. It is due to the prescribed nighttime heating emission profile used in the AirParif anthropogenic inventory (the second paragraph in section 3.2.1) and the nighttime model transport issue (the third paragraph in section 3.2.1).

Thirdly, the authors concluded that the transport issue is difficult to identify. I disagree. The model transport can be evaluated with meteorological observations. Apparently, there are meteorological observations at the monitoring sites. Additionally, there are quite a few WMO stations in the domain. Comparing with meteorological observations in the model domain will allow the authors to have a better sense of the model transport.

We certainly agree with the reviewer that the accuracy of the modeled CO_2 concentrations depends on the quality of the meteorological model. Therefore, as a first step, our previous Lian et al. (2018) paper particularly focuses on the evaluation of WRF in simulating the meteorological variables over the Paris region. The statistics for WRF results as compared to the observations show that our model setup for Paris with its multi-nudging options can well reproduce the near-surface air temperature and wind without obvious technical or configuration issues (Figure S2 in Lian et al. 2018). We thus feel that there is no need to fully resume such a detailed meteorological validation in this study. We acknowledge nevertheless that the reader may want to see more so that we have provided the following assessment of the meteorological fields.

We have added a new sub-section (section 3.1.1 Meteorological fields) together with the two new figures (Figure 4 and Figure S2) in the revised manuscript.

“In this section, we start with an evaluation of the overall performance of the control run (BEP_MYJ) in simulating both meteorological fields and atmospheric CO_2 over the full-year period. Since the accuracy of the modeled CO_2 concentrations depends on the quality of the meteorological model, the simulated meteorology by WRF was first evaluated against observations at SAC100 and SIRTA stations with a focus on three variables (air temperature, wind and PBL height). Figure 4 shows the time series of the 1-year daily afternoon mean (11-16 UTC) observed and modeled temperature, wind speed and wind direction at SAC100 station, together with their statistics summarized in the scatter plots. The daily nighttime mean (21-05 UTC) data are shown in Figure S2. In general, both daytime and nighttime temperature are well reproduced by WRF with correlation coefficient, RMSE and MBE of 1.0, 0.44°C, 0.06°C and 0.99, 0.67°C, 0.23°C respectively. The analysis of the MBE shows that the wind speeds are slightly overestimated by WRF, with a bias of 0.96 m/s for afternoon and 0.68 m/s at night. As for the wind direction, the model-data misfits decrease with the increasing wind speed. Seasonal (and even some day-to-day) variations in the afternoon average PBL heights diagnosed from the model data are in general agreement with the observations at the suburban SIRTA site with a RMSE of 359 m and a positive bias of 82 m. Some disagreements between the model-data PBL height estimates can be expected given layer heights from aerosol-based methods (as here applied to the observations) tend to lag behind those determined from thermodynamic methods (applied to the model data) during the course of the day (Kotthaus et al., 2018). Relative agreement between PBL heights is reduced at night (Figure S2), as uncertainties are higher in both the observed layer heights (Section 2.2) and those diagnosed from the model data (Shin and Hong, 2011). In general, results in Figure 4 and Figure S2 show that the simulated meteorological fields agree reasonably well with observations both during day and night which indicates parameter settings suitable overall.”

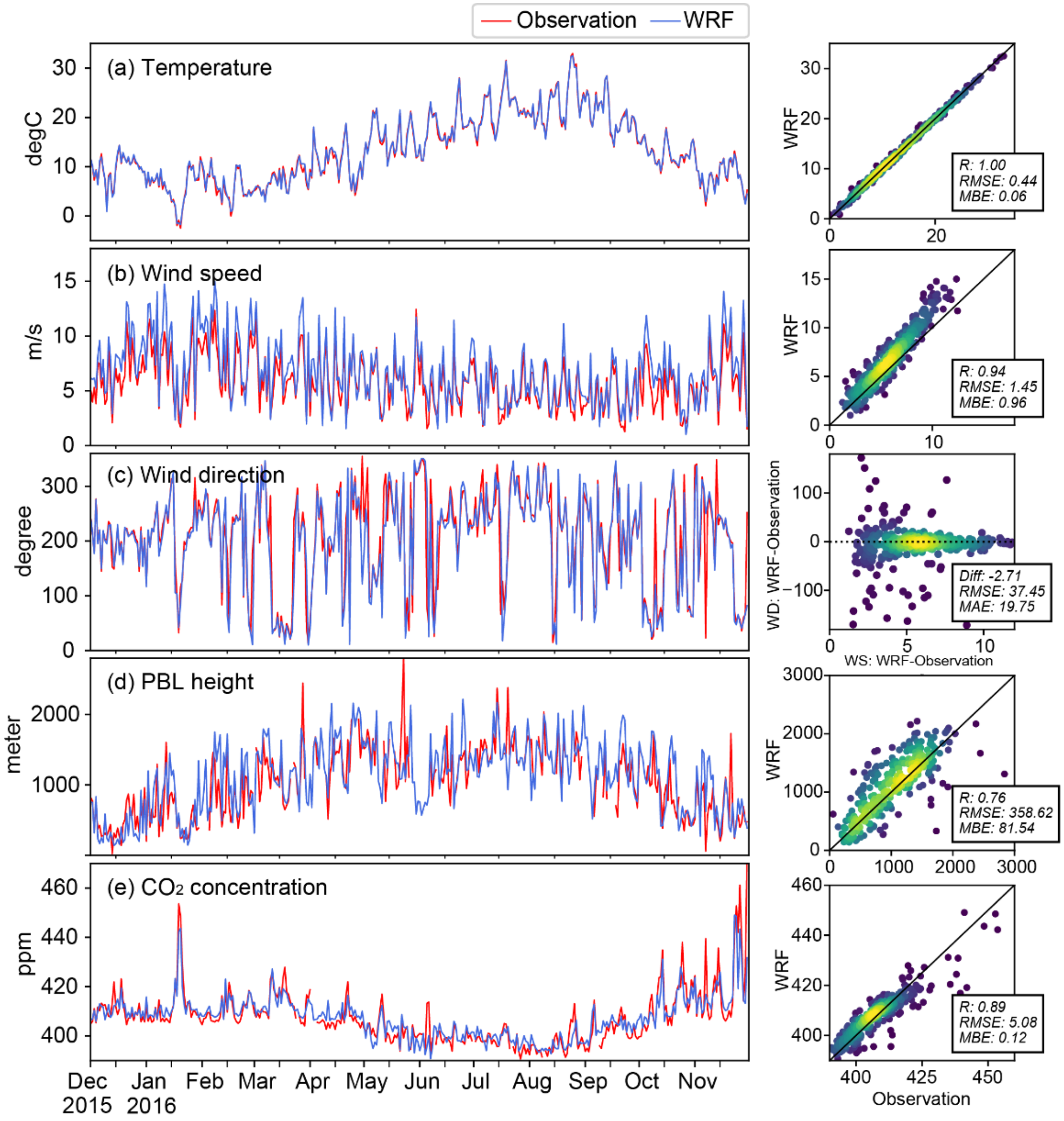

Figure 4. Time series of the daily afternoon mean (11-16 UTC) observed and BEP_MYJ modeled (a) temperature, (b) wind speed, (c) wind direction and (e) CO_2 concentration at SAC100 station. (d) Time series of the daily afternoon mean (11-16 UTC) observed and modeled PBL height at SIRTA station.

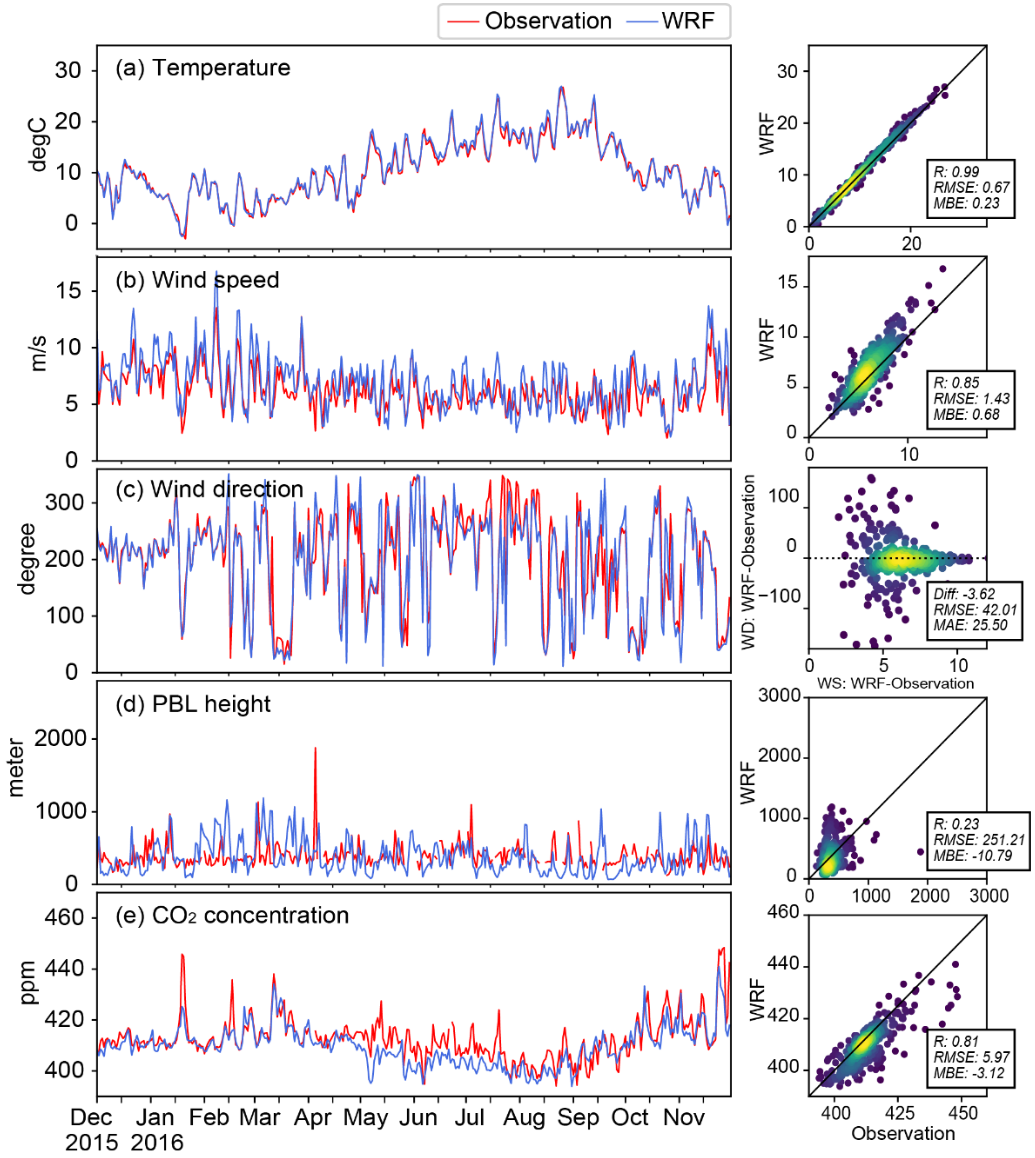

Figure S2. Time series of the daily nighttime mean (21-05 UTC) observed and BEP_MYJ modeled (a) temperature, (b) wind speed, (c) wind direction and (e) CO_2 concentration at SAC100 station. (d) Time series of the daily nighttime mean (21-05 UTC) observed and modeled PBL height at SIRTA station.

As the authors mentioned that boundary conditions can lead to large bias in the inversed results, the results showed that 5-20 ppm day-to-day difference between the two global models along the edges of the model domain. In the CO_2 regional (inverse) modeling, one of the major concerns about the boundary condition is the conservation of mass (Butler et al., 2020). How did the authors handle mass conservation when incorporating global modeled CO_2 into the regional model domain?

As described in the manuscript, both global datasets (CAMS and CarbonTracker) were interpolated onto the outermost domain of WRF-Chem (D01) (bilinearly in longitude, longitude and linearly in pressure) so as to provide the lateral boundary conditions for CO_2 simulations. We did not specifically address the pressure-weighted integrated columnar concentration of CO_2 as we only focus on the model-data comparison for the near surface in situ CO_2 measurements rather than the column-average dry-air mole fractions of CO_2 (XCO2) measured by the satellite or the ground-based remote sensing system. In addition, it is worth pointing out that west winds (180-360° headings) are dominant in the Paris area. For most time of the year (~73%), the differences in simulated CO_2 concentrations over Paris are within the range of ±1 ppm since they are mainly affected by those differences between CAMS and CarbonTracker at the western boundary of D01. Even though the differences between CAMS and CarbonTracker are larger at the eastern boundary (-4.8 ± 7.4 ppm for 00 UTC and -1.7 ± 3.3 ppm for 12 UTC), the magnitude of uncertainties becomes much smaller after a long-distance transport of CO_2 (up to 5 ppm during several synoptic episodes). Under these circumstances, we also suggest the use of CO_2 gradients between upwind-downwind stations in the inversion for Paris, which will further decrease the impacts as compared to a mass-balance inversion method.

Another issue is that the number of boundary conditions used is too small to quantify the uncertainty. Strictly speaking, to be able to claim quantification of the uncertainty sources, a large number of the ensemble and a set of calibration procedures are required, such as rank histogram, reliability diagram, brier scores, etc. (Garaud and Mallet, 2011). Although it may be difficult to meet two criteria with the CO_2 modeling, the authors will at least need three of them to study the sensitivity.

We disagree on this point. By using the two state of the art global CO_2 atmospheric inversion products (CAMS and CarbonTracker), the results in this study do show a fairly detailed information of uncertainties linked with the boundary condition hypothesis. It also provides an insight for the use of CO_2 gradients between upwind-downwind stations in the inversion of CO_2 fluxes for Paris so as to remove the potential errors from the boundary conditions, in particular when winds blowing from the east during the period of inversion. It is worth noting that the sensitivity tests in this study do not intend to explore and cover the full uncertainty space. We are more interested in the order of magnitude with the current two realistic boundary conditions rather than the theoretical perturbations with three members (or even more). With this perspective, using these two most suitable products is well enough to achieve our objective.

In summary, this work claims that it has a quantitative evaluation of uncertainty sources in the CO_2 modeling, but the experimental design is far from achieving the goal. It eventually is merely a sensitive study of modeled CO_2 to the selected fossil fuel emissions, the combination of PBL and urban canopy models, and boundary conditions. The size of the ensemble they built does not allow them to do a solid quantification study. As I mentioned, this study appears repeating some of the previous studies without advancing the understanding the community already holds currently, neither in science nor in techniques.

It is well known that city-scale inversion bears a large number of challenges that we do not claim to solve at once. We believe that the present manuscript provides error characterization and a range of new detailed insights on the signature of the different types of sources of errors at city scale, that most likely will remain during the forthcoming years. Meanwhile, following the suggestion from Reviewer #3, we have modified the title to better reflect our intent.

There are no clear rationales why they made such selection as I pointed out with the transport “ensemble”. The authors did not address the major issues in urban modeling, i.e., the impact of biosphere, and regional modeling, i.e., the conservation of mass when applying boundary conditions. They also failed to have a clear conclusion about the findings associated with fossil fuel emissions. In my opinion, this work is incomplete and must be extended to consider publication; these concerns I brought up can be addressed, which, however, will require a new design of the method. In addition to the specific comments I listed below, I would not recommend this MS to be a published in ACP.

Please see our answer above in terms of the VPRM biogenic flux, the conservation of mass, and the conclusion about the fossil fuel emissions.

Specific comments:

Section 2: There are important details missing in the description of the model setup.

1) Did the model use simulation cycles? If yes, how often is it? If yes, how was the CO_2 field addressed, initializing every time or being carried over simulation cycles?

Following the reviewer's suggestion, we have added the following sentence in section 2.1.3 to address this point:

"The simulation was restarted every 5 days with the CO_2 initial values from the previous run."

2) ERA-Interim and the outermost domain of WRF-Chem have quite different resolutions. What are the rationales that the authors used grid nudging over spectral nudging?

As mentioned above, the choice of the model configuration in this study corresponds to those selected by Lian et al. (2018) which showed good performances for a representation of meteorological fields in support of the atmospheric transport modeling. The impact of grid nudging, spectral nudging, and WRF OBSGRID program have been investigated in Lian et al. (2018). We used the combination of grid nudging, surface analysis nudging and observation nudging together with the objective analysis (the latter three are generated by OBSGRID) to maximize the benefit of assimilating surface and upper air meteorological observations. The model performance of this multi-nudging options (WRF_OA) was compared to the one with the spectral nudging (WRF_noOA). Results show that WRF_OA provides obvious improvements in modeled surface temperature and wind speed over WRF_noOA, and is therefore recommended to be used in the atmospheric transport modeling when an accurate description of reality is needed. More details regarding the nudging to different variables and their coefficients are all described in Lian et al. (2018).

We feel that there is no need to explain in detail why we used grid nudging instead of spectral nudging in this study for two reasons:

(1) Comparison of the performance of nudging techniques (e.g. grid nudging vs. spectral nudging) is not closely related to the objective of this study. It has already been investigated in our previous study and referred in this manuscript.

(2) The nudging strategy used in this study has already been described in section 2.1 as follows: "The grid nudging option in WRF to relax the model to ERA-Interim on large scales was applied to temperature and wind fields at model levels above the planetary boundary layer (PBL) of the outer two domains. We also used the surface analysis nudging and observation nudging options to assimilate the National Centers for Environmental Prediction (NCEP) operational global upper-air (ds351.0) and surface (ds461.0) observation weather station data (https://rda.ucar.edu/datasets/ds351.0/; https://rda.ucar.edu/datasets/ds461.0/), which are described in more detail in Lian et al. (2018)."

3) As I mentioned in the general comments, have the VPRM parameters constrained with the flux tower measurements?

See answer above. In this study, the VPRM parameters have already been optimized with the eddy covariance flux measurements over Europe.

4) When the authors were incorporating CO_2 IC/BC to WRF-Chem, how did the author address the conservation of CO_2 mass?

See answer above.

5) When using global modeled CO_2 as IC, the discontinuity of the global and regional model dynamic can cause discrepancy of the CO_2 as well. How much the difference caused by the discontinuity would be?

We follow the traditional method used in the modeling community for the IC & BC interpolation and the WRF downscaling. The WRF outermost boundaries were set far away from the area of our interest, with three levels of nesting with horizontal grid spacing of 25, 5 and 1 km, covering Europe (D01), Northern France (D02) and the IdF region (D03) respectively. We thus believe that such a discontinuity of the global and regional model dynamic at the D01 boundaries is not a critical issue for the simulated CO_2 concentration over Paris, especially for the CO_2 gradients across the city. This paper did not aim at solving such specific secondary problems in depth, which is out of the scope of this study.

P 6, L 25-30: the author interoperated that the reason of the higher CO_2 concentrations in fall than in winter was due to the anticyclone keeping the high CO_2 in the domain for quite a while. I disagree. If it’s due to meteorology, the impact on the fossil fuel CO_2 and biospheric CO_2 concentration should be the same. We should see lower CO_2 in the suburban sites, but we don’t.

The figure below (Figure R2) and analysis of weather regimes show that the higher concentration in autumn 2016 are partly due to more anticyclonic events associated with CO_2 peaks from mid-October to mid-November 2016 compared to December 2015. The impact on the concentrations is not only a function of the atmospheric circulation, but also on the fluxes. Tracer simulations (Figure R2c and R2d) indicate that these peaks in autumn 2016 are mainly explained by a higher contribution of anthropogenic emissions compared to those of 2015, whereas the contribution of biogenic respiration is similar between autumn 2016 and winter 2015.

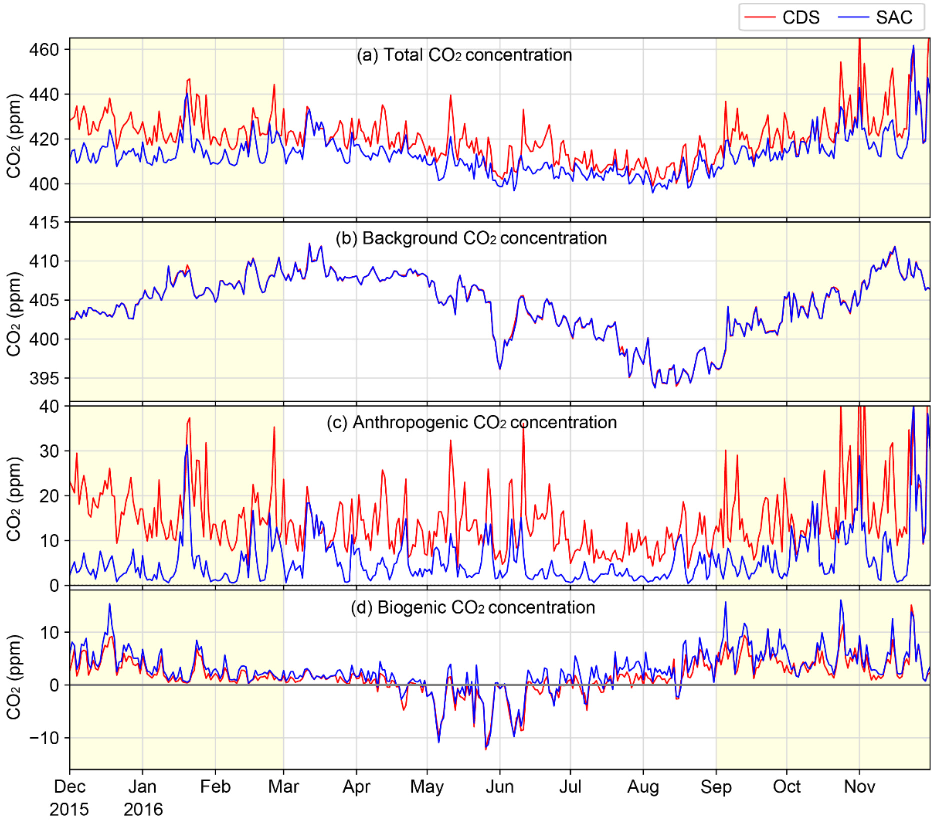

Figure R2. Time series of the BEP_MYJ simulated daily average (a) total, (b) background, (c) anthropogenic and (d) biogenic CO_2 concentration at CDS and SAC station

P7, L1-5: as I said earlier, the authors should be able to identify at least to some degree if the issues are in transport or boundary conditions by comparing with the meteo data.

See answer above.

P10, L10-15: what causes the different bias between the BEP and UCM schemes? I would like to see a deeper explanation of that instead of simply saying lower or higher.

Concisely, the two schemes differ in their representations of the near-surface mixing which leads to large differences in the modeled CO_2 concentrations. Figure S7a shows the annual average of the vertical distribution of CO_2 concentrations at JUS station for 24 hours a day for the two schemes BEP, UCM and their differences. It can be seen that both BEP and UCM schemes reproduce large vertical gradients in CO_2 concentrations in the low atmosphere levels, i.e. up to approximately 300 m AGL but mostly in the first 100 m. In general, the BEP scheme reproduces smaller CO_2 concentrations in the lower part of the atmosphere (<100m) than UCM does. After the sunrise at 5 UTC, the mixing by turbulent diffusion and afterwards by thermal plume dilute CO_2 to higher levels more quickly in BEP than UCM. The height of the boundary layer keeps increasing between sunrise and noon. During the afternoon, the boundary layer is well developed with enough mixing, resulting in CO_2 being transported to the upper layers through atmospheric convection. Figure S7b shows the vertical distributions of CO_2 concentrations during the afternoon (11-16 UTC) at JUS station for 12 calendar months. The UCM scheme reproduces a much larger vertical gradient in CO_2 concentrations close to the surface than the BEP scheme does, especially in the winter time. This is because of the high emissions (mainly household heating) and the more stable atmosphere in winter than in summer.

More precisely, the different depictions of the urban canopy parameters in the two modules (e.g. building heights, pervious area fractions, street canyons, heat capacity and thermal conductivity) and their impact on the energy budget and atmospheric transport are most likely to be the cause of the performance difference between BEP and UCM. The simulated near-surface CO_2 concentrations are highly sensitive to small differences in urban friction coefficient, wind velocity and vertical mixing associated with turbulence over the emission-rich areas (e.g. city center in winter). Even though we have modified the geometric and thermal parameters in the module over Paris based on the work of Kim et al. (2013), there are still many land surface characteristics that were not calibrated due to lack of detailed information of the urban form for the Paris city. A further improvement of the urban canopy schemes and a deeper analysis are out of the scope of this study.

Based on the discussion above, we have added the following sentence in the manuscript and Figure S7 in the supplement to account for the reviewer’s comment:

“This is because the BEP scheme generates more mixing in the lowest atmosphere especially from 07 to 14 UTC in the day and in winter relative to summer, which reduces the vertical gradient and therefore the largest concentrations near the surface (Figure S7).”

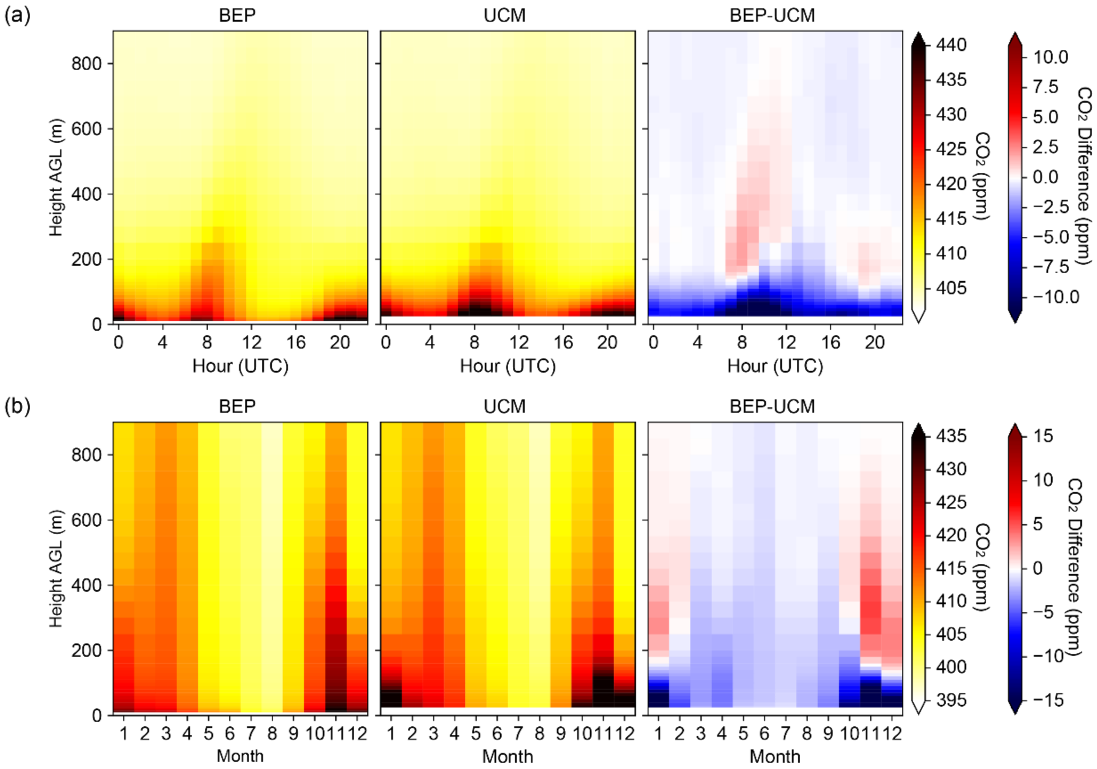

Figure S7. (a) Annual average of the vertical distributions of CO$_2$ concentrations at JUS station for 24 hours of the day for BEP, UCM and their differences; (b) Vertical distributions of CO$_2$ concentrations during afternoon (11-16 UTC) at JUS station for 12 calendar months for BEP, UCM and their differences.

P11, L19-21: I agree that based on the current setup, there is little hope to improve the model performance. However, the authors can follow my suggestion listed in the general comments. For example, checking the land surface model used, comparing with meteo data, etc., to identify if the problem is caused by transport is the first step. Then the authors can look into the emission, boundary conditions, etc.

Please see our answers above in terms of the choice of physics schemes, the model-data validation for the meteorological variables. Our previous study and the subsequent analyses (section 3.1.1 together with Figure 4 and S2) in this study indicate that the model-data discrepancy is not merely due to an obvious error in the atmospheric transport modeling. We thus analyze in further detail these measurement-model discrepancies and attempt to identify cases when they appear to be mainly driven by uncertainties in the anthropogenic emissions, in the biogenic fluxes, in the physical parameterizations of the atmospheric transport model, or in the CO$_2$ boundary conditions, as presented in section 3.2.

Figure 5: please use local time in the x-axis instead of UTC. The much bigger issue is the large bias in the biases.

Note: Figure 5 is now ranked as Figure 6 in the revised manuscript.

The local time in Paris is one hour ahead of UTC (UTC+1) from November to March, and two hours ahead of UTC (UTC+2) from April to October. As the time zone difference is only one or two hours from UTC to the local time, it may not seriously affect our visual interpretation of the results. Moreover, given that the time scale for the other figures and text in this manuscript are all shown as UTC, we might prefer to use UTC in the x-axis of Figure 6. We have added the following text in the caption of Figure 2 to clarify this point:

“The local time in Paris is one hour ahead of UTC (UTC+1) from November to March, and two hours ahead of UTC (UTC+2) from April to October.”

Please see our answer above in terms of the large bias.

The following references are used only in the responses to the reviewer’s comments. All other references mentioned above are already included in the manuscript.

Kim, Y., Sartelet, K., Raut, J. C., and Chazette, P.: Evaluation of the Weather Research and Forecast/urban model over Greater Paris. Boundary-layer meteorology, 149(1): 105-132, 2013.

Lac, C., Donnelly, R. P., Masson, V., Pal, S., Riette, S., Donier, S., Queguiner, S., Tanguy, G., Ammoura, L., and Xueref-Remy, I.: CO_2 dispersion modelling over Paris region within the CO2-MEGAPARIS project, Atmospheric Chemistry and Physics, 13, 4941-4961, 2013.

Reference:

Butler, Martha P., Thomas Lauvaux, Sha Feng, Junjie Liu, Kevin W. Bowman, and Kenneth J. Davis. “Atmospheric Simulations of Total Column CO_2 Mole Fractions from Global to Mesoscale within the Carbon Monitoring System Flux Inversion Framework.” Atmosphere 11, no. 8 (August 2020): 787. https://doi.org/10.3390/atmos11080787.

Díaz-Isaac, Liza I., Thomas Lauvaux, and Kenneth J. Davis. “Impact of Physical Parameterizations and Initial Conditions on Simulated Atmospheric Transport and CO_2 Mole Fractions in the US Midwest.” Atmospheric Chemistry and Physics 18, no. 20 (October 16, 2018): 14813–35. https://doi.org/10.5194/acp-18-14813-2018.

Feng, S., Lauvaux, T., Newman, S., Rao, P., Ahmadov, R., Deng, A., et al. (2016). Los Angeles megacity: a high-resolution land–atmosphere modelling system for urban CO_2 emissions. Atmospheric Chemistry and Physics, 16(14), 9019–9045. https://doi.org/10.5194/acp-16-9019-2016

Feng, Sha, Thomas Lauvaux, Kenneth J. Davis, Klaus Keller, Yu Zhou, Christopher Williams, Andrew E. Schuh, Junjie Liu, and Ian Baker. “Seasonal Characteristics of Model Uncertainties From Biogenic Fluxes, Transport, and Large-Scale Boundary Inflow in Atmospheric CO_2 Simulations Over North America.” Journal of Geophysical Research: Atmospheres 124, no. 24 (2019): 14325–46. https://doi.org/10.1029/2019JD031165.

Feng, Sha, Thomas Lauvaux, Klaus Keller, Kenneth J. Davis, Peter Rayner, Tomohiro Oda, and Kevin R. Gurney. “A Road Map for Improving the Treatment of Uncertainties in High Resolution Regional Carbon Flux Inverse Estimates.” Geophysical Research Letters 46, no. 22 (2019): 13461–69. https://doi.org/10.1029/2019GL082987.

Garaud, D., and V. Mallet. “Automatic Calibration of an Ensemble for Uncertainty Estimation and Probabilistic Forecast: Application to Air Quality.” Journal of Geophysical Research: Atmospheres 116, no. D19 (October 16, 2011). https://doi.org/10.1029/2011JD015780.

Hilton, T.W., K. J. Davis, and K. Keller. 2014. Evaluating terrestrial CO_2 flux diagnoses and uncertainties from a simple land surface model and its residuals, Biogeosciences, 11, 217-235, doi:10.5194/bg-11-217-2014.

Hilton, T.W., K. J. Davis, K. Keller, and N.M. Urban. 2013. Improving North American terrestrial CO_2 flux diagnosis using spatial structure in land surface model residuals, Biogeosciences, 10,4607–4625, doi:10.5194/bg-10-4607-2013.