Anonymous Referee #2

This paper analyzes the potential of the WRF-GHG model to simulate the column-averaged abundances of CO_2 and CH_4 under urban environments. The precision of such simulations is assessed by comparing to observations carried out in Berlin using portable and low-resolution EM27 spectrometers. The combination of model and observations allow the authors to highlight the potential of using the differential column methodology (DFM) to estimate urban CO_2 and CH_4 emissions and identify the main driver processes responsible for such emissions. The paper is well structured, but extremely concise and the results are not properly discussed in depth. I suggest this paper might be suitable for publication after general and specific comments listed below are addressed.

We thank the anonymous Referee #2 for their time and valuable comments to improve this manuscript. The general and specific comments are addressed by point-by-point detailed replies below. Referee's comments have been repeated in black. Author's replies and edited contents are marked in blue and red, respectively. We would like to add Frank Hase and Matthias Frey as the co-authors in this paper because of their contributions in the measurements.

General comments

1) The useful and potential of the differential column methodology (DFM) is investigated to evaluate the urban emissions. It is especially interesting the author's suggestion to eliminate the wind influence on this methodology. But, in general, this promising database (simulations and observations) is poorly exploited by the authors. The discussion and attribution of GHGs sources is very simple, not being supported by robust statistics methods (eg, multivariate analysis,...) and, in some parts of paper, only using one day of observations. Also, I would recommend to the authors to go in depth in this work by providing flux emission estimations from this database, similarly for example to Viatte et al. (2017). Specially, from the observation point of view, it is crucial to analyze the potential of DFM technique to provide actual GHGs flux emissions. Other interesting point would be to assess the improvements of WRF-GHG simulations if the atmospheric GHGs observations from EM27 instruments were assimilated

Response: Thanks for your valuable comments. The focus of this study is to evaluate the behavior of WRF-GHG in urban areas and test whether differential column methodology could potentially be a proper method in the model analysis to cancel out the bias from background concentrations and highlight the major emission tracers within limited regions. Emission flux estimations using WRF-GHG would be our further target which plans to be demonstrated in the Munich case. This Munich case is combined with the first worldwide permanent column measurement network designed in Munich. Various emission tracers are suggested to run for this case in which more emission tracers (e.g., biogenic emissions from wetland for XCH_4, the traffic emission and strong point sources' emissions in urban areas) are being added and the long-period data can also be available. Thus, this Berlin case is a fundamental study for the WRF-GHG model with high resolutions in urban areas and testing whether differential column methodology can work for the model analysis in urban regions. The reason why we only consider two dates for concentration comparison at the beginning is the limitation of EM27 measurement status and we chose to follow Hase et al.,(2015) and target the two best measurement dates. In the rephrased version, all the dates with the relatively good measurement qualities (above 'fair') are shown. And we also included the calculation of the smoothed simulations using the EM27 vertical sensitivity and compare the smoothed values to the values without smoothing which are described in Sec.3.3 and Appendix.D.

2) The precision of the WRF-GHG simulations is assessed by comparing to the EM27 column observations carried out between 23rd June and 11th July 2014 in Berlin. But, the authors use different days depending on what they want to discuss without a precise explanation. For example, the wind comparison is only performed for the period from 1st July to 5th July, the comparison of column-averaged concentrations is carried out for the 3rd and 4th July. But, for the DCM analysis the authors include the 1st and 2nd July, but they rule out the 4th July for section 4.1. However, for section 4.2 the authors use 3rd, 4th, 5th, and 6th July. This limitation on dates also affects the robust of the found conclusions, since the size of the compared database is very limited (for example, the section 4.2 is almost supported by the comparison of one day, analyzing correlations with only ten points). This is very confusing. Why do not the authors use the whole field campaign observations for the wind and concentration comparison? As authors pointed out, to check the DCM approach stable conditions are needed, but these conditions must be justified and supported by the results from this work itself. Therefore, I would recommend to authors to include the whole period of the field campaign for the comparison of the wind and column-averaged concentration fields (sections 2 and 3), which will also support the identification of GHGs

sources and sinks presented in the paper. From these comparison results, the authors could clearly identify the optimal conditions to analyze the DCM approach.

Response: Thank you for this comment. We had shown limited (representative) time periods in order to improve clarity, but we agree with the referee that our choice may not have been optimum. So in the rephrased manuscript, we took all the dates when the measurement qualities are above 'fair' into account (basically, 5 simulation dates, 1st, 3rd, 4th, 6th and 10th July).

Specific comments

Line 74: The novelties and obtained improvements with regards to the work of Pillai et al. (2016) should be discussed in more detail in the paper.

Response: We have included in Line 77-85:

Pillai et al. (2016) utilized a Bayesian inversion approach based on WRF-GHG at a high spatial resolution of 10 km for Berlin to obtain anthropogenic CO_2 emissions, and to quantify the uncertainties in retrieved anthropogenic emissions related to instruments (e.g. CarbonSat) and modelling errors. In the present paper, our focus is on a high-resolution (1 km) study of both CO_2 and CH_4 in Berlin, and assess the performance of WRF-GHG through comparing the simulated wind and concentration fields to observations from wind stations and ground-based solar-viewing spectrometers. Then DCM is tested as a proper approach for model analysis, which can cancel out the bias from initialization conditions and highlight regional emission tracers. The simulation workflow is also adapted to this purpose where needed. This study is the fundamental study of the WRF-GHG mesoscale modeling framework in urban areas.

Line 125: One of the novelties of this work is to provide GHGs high-resolution simulations (1km x 1km). In this sense, it would be nice to discuss the impact of using, for example the EDGAR V4.1 inventory as anthropogenic fluxes at a spatial resolution of 0.1 (about 10 km x 10 km), on the simulated concentrations fields, or how the spatial resolution is treated in the VPRM model.

Response: Thank you for the valuable comment. The WRF model with finer resolutions is able to capture more details of local modifications to the simulated fields (e.g., wind fields (DuVivier, A. K., & Cassano, J. J., 2013)) and perform better than the coarse-resolution model for all simulations (Jee, J. B., & Kim, S., 2017). Compared to the coarse-resolution model, the fine-resolution WRF-GHG model in this study should be able to capture better meteorological fields and simulation fields which are suitable for comparing with the ground-based measurements.

In terms of the anthropogenic tracer, we use a quite simple method to downscale EDGAR inventory from 10 km to 1 km. Basically, each model grid is matched with the closest grid point in EDGAR inventory and saving the corresponding EDGAR emission flux to the initialization file of our model. During the simulation, this initialization file is read into the first layer of our domain. Then, combined with the meteorological conditions and boundary conditions for concentration fields, the data from this initialization file are used for the 'online' calculation, i.e. the WRF-GHG model.

For the biogenic tracer, in turn, VPRM calculates the hourly NEE based on MODIS satellite estimates of LSWI, EVI, etc. (Beck et al., 2011). Before using vegetation indices in WRF-GHG initialization, reflectance data from the MODIS satellite (8-day intervals and 500 m spatial resolution) is combined with Synmap vegetation classification (1 km resolution) and per vegetation class transformed to the desired projection (Lambert Conical Cartesian co-ordinate system in our case) within the WRF-VPRM preprocessor. A brief information on the treatment of anthropogenic and biogenic tracers has been included in Line 137-142:

The biogenic CO_2 emission is calculated online using VPRM (Mahadevan et al., 2008), in which the hourly Net Ecosystem Exchange (NEE) of CO_2 reflects the biospheric fluxes between the terrestrial biosphere and the atmosphere, estimated by the sum of Gross Ecosystem Exchange (GEE) and Respiration. VPRM in WRF-GHG calculates biogenic fluxes initialized by vegetation indices (land surface water index (LSWI), enhanced vegetation index (EVI), etc..) from the MODIS satellite (as available via https://modis.gsfc.nasa.gov/). Combined with SYNMAP vegetation classification at a resolution of 1 km, the refectance data from the MODIS satellite at a 500-m spatial resolution and 8-day intervals, is aggregated to the Lambert Conformal Conic (LCC) projection within the WRF-VPRM preprocessor. Then, the data including these high-solution vegetation indexes at a resolution of 1 km are available on the model domains.

Line 129: The authors mention that the time factors for the assimilated anthropogenic fluxes could introduce considerable uncertainties. Please, explain more in detail this issue, eg, the value of these uncertainties or how these could affect the GHGs concentration simulations.

Response: As documented in the EDGAR official website (https://themasites.pbl.nl/tridion/en/themasites/edgar/documentation/content/Temporal-variation.html), the temporal-variation set is originally made by Veldt (for the chemistry transport model (CTM) Long Term Ozone Simulation (LOTOS), a European climate model) and contains the variations for seasons (winter/summer), day-to-week (weekday/weekend) and day-and-night. This temporal variation set is defined based on Western European data. Thus, uncertainty can arise when applying this set to other European countries and even more so for other regions. Furthermore, the coarse anthropogenic-emission initialization in our study makes it hard to distinguish and locate the specific emission points within high-resolution model grids and weakens the impact from the real high emission hot-spot in urban area. We add information about these caveats in Line 149-153:

Here we apply time factors for seasonal, daily and diurnal variations defined by the time profiles published on the EDGAR website (http://themasites.pbl.nl/tridion/en/themasites/edgar/documentation/content/Temporal-variation.html); however, considerable uncertainties are to be expected in applying these time factors. This temporal variation set is derived based on western European data such that the representativity for other European countries and even other world regions may be quite poor. The coarse emission fluxes used for the initialization of the anthropogenic tracer in WRF-GHG can cause problems when locating emission points within the high-resolution model grid, and can weaken the impact from the real high emission hot-spots in the fine domain of our study.

Line 152. The field campaign is carried out between 23rd June and 11th July 2014. Why is the period from 1st July to 5th July only used for the wind comparison? Since EM27 observations are performed during day-time, it would be interesting to include the performance analysis for wind comparison distinguishing between day and nighttime.

Response: The simulation period for this Berlin case is from 1st July to 10th July. The figure which contains the comparisons of wind fields for ten days is too crowded. Figure 2 is updated already, and now includes wind comparisons for ten days. The reason why we do not include the period in June as our simulation period is that the measurement situations were not stable in June. Although the measurement campaign started from 23rd June, only the measured data for these 3 days (26th, 27th and 28th June) is provided (Hase et al., 2015) and the measurement qualities are not good (only 27th June is 'good', others are very poor). To be a fundamental study of the model framework which can run continuously and comparing the simulations with a relatively stable measurements, we decide to start our simulation run from the start of July.

We highlight the wind fields for the measurement periods (basically in the day time between 7 am to 16 pm), using the gray-shadowed squares in Fig.2 of the content. RMSE values are used to show the performance of wind fields. We conclude that the simulations fit better with the measurements in the day time, compared to the values in the nighttime. The measured wind fields in the nighttime holds larger variability compared with the simulation. All these contents are in Line 176-190:

Figure.2 shows the comparisons of wind speeds (Fig.2(a)) and wind directions (Fig.2(b)) between simulations and observations at 10 meters from 1st July to 10th July and the model-measurement differences. EM27/SUN only operates in the daytime under enough sunlight (the detailed description of the instrument can be found in Gisi et al.(2012), Frey et al.(2015) and Vogel et al.(2019)). The instrumental working periods are marked by gray shaded boxes in Fig.2. The measured (dashed lines) and simulated (solid) wind speeds (Fig.2(a)) at 10 meters show similar trends and demonstrate relatively good agreement over the 10-day time series with a root mean square error (RMSE) of $0.9247\,\mathrm{m/s}$. Large uncertainties in wind speeds are found to appear always with the lower wind speeds, mostly at night. In terms of wind directions at 10 meters, we observe that the simulated wind directions show similar but slightly underestimated fluctuations (Fig.2(b)), which result in a RMSE of $60.8328°$. Larger uncertainties in wind directions always exist during the low wind speed periods (Fig.2(a)&(b)). During the instrumental working period (within the daytime), the simulations fit better with the measurements with relatively lower RMSEs $0.6928\,\mathrm{m/s}$ for wind speeds and $41.4793°$ for wind directions. We find that the measured wind fields (both wind speeds and wind directions) have more fluctuations, compared to the simulations. This could be caused by real fast wind changes, which the model, simulating a somewhat idealized environment, is not able to capture. To be specific, local turbulence given by urban canopy, buildings etc. are not represented well in the model.

Line 159. Explain possible reasons for the discrepancies between the model and observed wind fields. Maybe the model

is not able to capture very fast changes of air masses, which will strongly affect the GHG comparisons.

Response: Thanks for your comment. The possible reason for the discrepancies between the simulated and measured wind fields have been added in the content (line 187-190).

Line 166. The influence of the EM27 vertical sensitivity is neglected in the model-observation comparison. But, the EM27 averaging kernels show a dependence on the solar geometry, which is considered in similar studies such as Vogel et al. (2018). Please justify more in detail why not to smooth the GHGs simulations (using the EM27 averaging kernels and a priori profile).

Response: Really thanks for your suggestion and we have added the smoothing part in the content (in Sect.3.3). We do agree with your opinion and smooth the simulation results, using the EM27 column sensitivities associated with solar zenith angels (SZA) and the a-priori CO_2 profile provided by the Whole Atmosphere Community Climate Model (WACCM) Version 6. Within such a limited simulation period in July (10 days) in our study, the daily variations of SZAs for these ten days are mostly overlapped (Fig.1(a) of this response). That is to say, we can use the simulation date with the most measurement data (4^{th} July) as the reference for the column sensitivities with different SZAs. Then the column sensitivities following the model vertical pressure axes can be derived through interpolation on the basic of the reference column sensitivities.

As seen in Fig.1(b) and (c) of this response, the interpolated column sensitivities (the circles) fit well with the distribution of column sensitivities with different SZAs (the solid lines) for both CH_4 and CO_2. Compared to Vogel et al.(2018), we choose to interpolate the column sensitivities directly, instead of calculating the column sensitivities based on formulas, because the best measurement date in Berlin campaign (4^{th} July) can provide us enough values to capture the vertical distribution features of column sensitivities accurately. The a-priori CO_2 and CH_4 profiles are provided by the Whole Atmosphere Community Climate model (WACCM) Version 6. The calculation details about the smoothed profiles for XCO_2 and XCH_4 have been rephrased in Sect.3.3 of the content and one example of this smoothing calculation is depicted in Fig.2 of this response for XCH_4 to better understand the smoothing process:

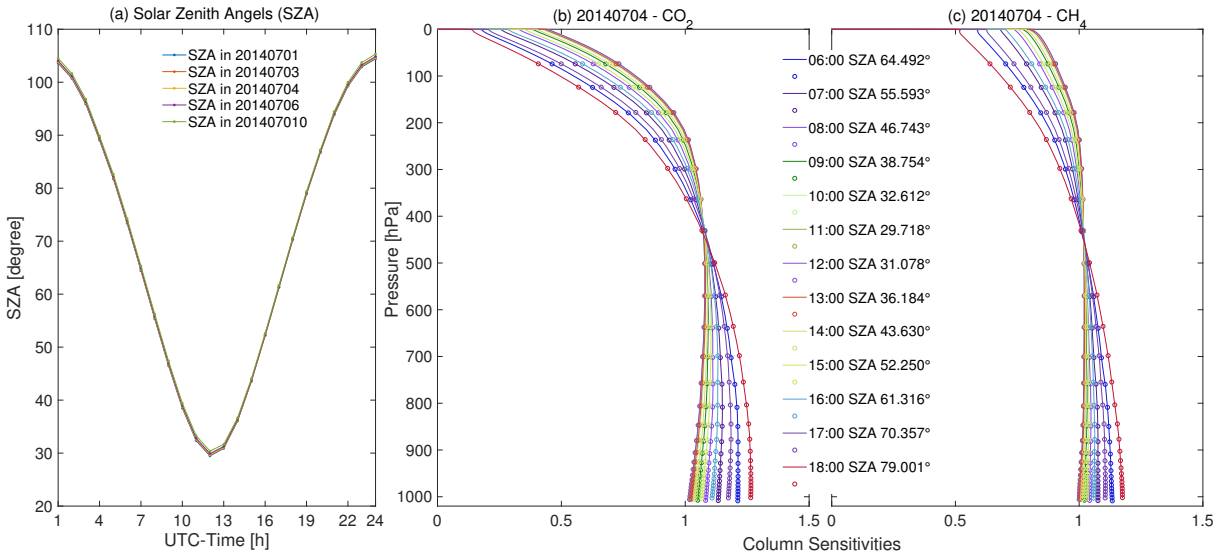

Figure 1: (a) The daily variations of solar zenith angels (SZA) for five simulation dates (1^{st}, 3^{rd}, 4^{th}, 6^{th} and 10^{th} July) and the vertical distributions of column sensitivities for (b) CO_2 and (c) CH_4 on 4^{th} July. (b) & (c): the solid lines represent the column sensitivities derived from EM27/SUN under different SZAs and the circles stand for the interpolated SZAs associated with the model pressure axes.

When comparing remote sensing observations to model data (or also datasets from different remote sensing instruments to one another), limitations of the instruments in reconstructing the actual atmospheric state need to be taken into account. In general, this requires the a-priori profile which was used for the retrieval and the averaging kernel matrix, which specifies the loss of vertical resolution (fine vertical details of the actual trace gas profile cannot be resolved) and limited sensitivity (e.g. Rodgers and Conner (2003)). In the case of EM27/SUN, the spectrometers used be the network offer

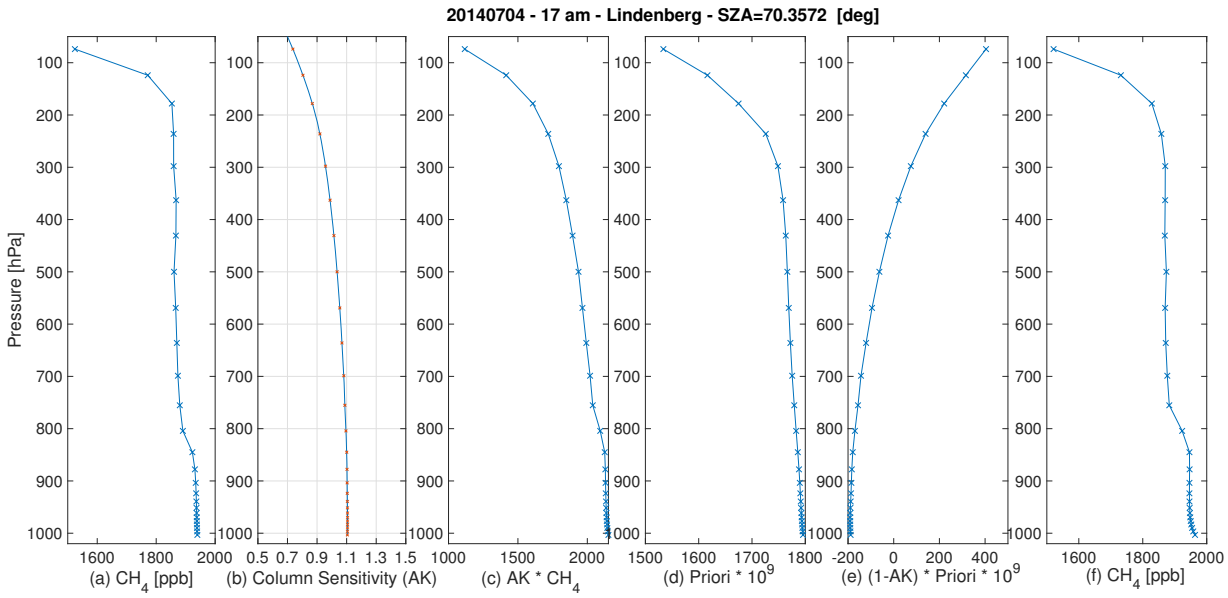

Figure 2: The calculations during the smoothing process. (a) the vertical concentration distribution derived from WRF-GHG output at 17 am on 4th July; (b) the solid line represents the column sensitivities derived from EM27/SUN and the orange dots stand for the interpolated column sensitivities associated with the model pressure axe; (c) the vertical concentration distributions weighted by the interpolated column sensitivities shown in (b); (d) the vertical distribution of concentrations derived from the priori model (WACCC); (e) the vertical distributions of the weighted (weight factors: 1-AK) priori concentrations; (f) the smoothed vertical concentrations profile obtained by the sum of (c) and (e).

only a low spectral resolution of $0.5\,\mathrm{cm}^{-1}$. Therefore, performing a simple least squares fit by scaling retrieval of the a-priori profile is generally appropriate. In this case, there is no need to specify a full averaging kernel matrix, instead, the specification of a total column sensitivity is sufficient. The total column sensitivity is a vector (being a function of altitude), which specifies to which degree an excess partial column superimposed on the actual profile at a certain input altitude is reflected in the retrieved total column amount. This sensitivity vector is a function of solar zenith angle (and ground pressure), mainly due to the fact that the observed signal levels in different channels building the spectral scene used for the retrieval are shaped by a mixture of weaker and stronger absorptions (if all spectral lines in the spectral scene would be optically thin and too narrow to be resolved by the spectral measurement, the sensitivity would approach unity throughout).

In order to ensure measurement qualities and enough sample points for further concentration comparisons, we select five measurement dates (1st, 3rd, 4th, 6th and 10th July) with relatively good measurement qualities (from fair '++' to very good '++++') based on Hase et al. (2015). The pressure-dependent column sensitivities for CO_2 (Fig.3(b)) and CH_4 (Fig.3(c)) are derived from measurements performed in Lindenberg on 4th July (the best measurement-quality day). Details about the measurements can be found in (Hase et al. (2015) and Frey et al. (2015). The shape and values of the column sensitivities from Karlsruhe closely resemble the results of Hedelius et al. (2016) in Pasadena. As depicted in Fig.3(a), the solar zenith angles (SZAs) are almost identical for each day in our study (at each hour), rendering the shape of column sensitivities (at a specific hour of the day) practically independent of the measurement date. The column sensitivities for 4th July (Fig.3(b,c)), are taken as a basis for our smoothing process below. The a-priori CO_2 and CH_4 profiles have been taken from the Whole Atmosphere Community Climate model (WACCM) Version 6 here. A smoothed profile for a target gas G is then obtained as Eq.3 in (cf. Vogel et al., 2019),

$$G^s = K * G + (I - K) * G^p \tag{1}$$

where G is the modelled profile from WRF-GHG, I is the identity matrix, K is a diagonal matrix containing the averaging kernel, and G^p is the a-priori profile.

In order to compare the simulated smoothed concentration fields with the observations, the simulated smoothed pressure-

weighted column-averaged concentration for a target gas G (XG) is calculated as,

$$\Delta p(i) = \frac{P(i) - P(i+1)}{P_{\text{sf}} - P_{\text{top}}} \rightarrow XG = \sum_{i=1}^{n} \Delta p(i) \times G^s(i) \tag{2}$$

Here, Δp_i is the proportional to the differences of the pressure values $P(i)$ at the bottom and $P(i+1)$ at the top of the i^{th} vertical grid cell; P_{top} and P_{sf} represent the hydrostatic pressures at the top and at the surface of the model domain, and $G^s(i)$ stands for the simulated concentration of the target gas G at the i^{th} vertical level.

In Figures.D1 and D2 of Appendix D, we compare the simulated XCO_2 and XCH_4 with and without smoothing. The simulated concentrations are only slightly enlarged after smoothing, approximately 1-2 ppm for XCO_2 and 2 ppb for XCH_4, while the variations are mostly not changed. Compare to the period with lower SZAs (at noon), the smoothed values in the morning and afternoon with higher SZAs hold relatively larger enlargements.

Line 176. For the comparison between simulated and observed column-averaged concentrations, the authors limit the analysis to the 3rd and 4th July. But, according to the wind comparison, the 3rd July shows the highest discrepancies between model and simulations. Therefore, observations and simulations could account for different air masses with different GHGs concentrations and influenced by different sources, which should be considered in the discussion. To account for this, the authors could plot the simulated-observed differences as a function of the wind differences. Although the EM27 stations do not coincide with the meteorological stations, figure 2 shows that the study area is very homogeneous with regards to wind fields. Thereby, these differences could be a proxy of the model inconsistences.

Adjust better the scales of figure 3 to clearer see the dispersion and distribution of the data, especially for 4th. Now, it is hard to identify what database shows more variability, which could be interesting to know if the simulations are underestimating or overestimating the real CH4 variability.

Response: Thanks for your helpful suggestion! We have shown the concentrations for five targeted simulation dates and enlarged CH_4 concentrations to figure out the trends (details see Fig.3 and Sect.3.3). All the contents for the comparison of CO_2 and CH_4 have been somewhat re-worked. Now it can be seen that the measured XCH_4 holds larger variations. We also tried to figure out whether the bias is associated with wind differences, but there is no obvious connection between them.

Figure.4(a) shows the measured and modelled variations of XCO_2 and XCH_4 for these five days. Compared to the measurements, the smoothed simulated pressure-weighted column-averaged concentrations for CO_2 (XCO_2) show quite similar trends but with approx. 1-2 ppm bias, indicated by a RMSE of $1.2534\,\text{ppm}$. The simulated XCO_2 are overestimated for 1st, 3rd and 4th July while on 6th and 10th July, the model is underestimated, which could attribute to the uncertainties from the coarse anthropogenic surface emission fluxes, background concentrations from CAMS (Sembhi et al., 2015), and the ignorance of the influence from the line of the sun sight.

Figure.4(b) shows the comparison of the pressure-weighted column-averaged concentrations for CH_4 (XCH_4) between observations and smoothed simulations on the five selected dates (1st, 3rd, 4th, 6th and 10th July). We find that there is an approximate offset of 50-60 ppb between observations and models (RMSE = $58.1082\,\text{ppb}$). The simulated XCH_4 is around 1860 ppb while the measured value is around 1810 ppb which is comparable to the values (1790-1810 ppb) observed at two Total Carbon Column Observing Network (TCCON) measurement sites in June and July 2014, Bremen in Germany (Notholt et al., 2014) and Bialystok in Poland (Deutscher et al., 2014). This bias of the simulated XCH_4 seems to be constant (around 2.7 %) each day. Thus, we introduce an offset applied to all sites for each simulation date to compare the model and the measured data, effectively removing the bias, which we attribute to a too high background XCH_4. The daily offset is assumed to be the difference between the simulated and measured daily mean XCH_4. After applying the daily offset, the measured XCH_4 shows a somewhat better agreement and a similar trend but with larger variability, compared to the simulation (RMSE = $3.1690\,\text{ppb}$). The smaller variations from the simulation results can, e.g., be caused by the error from the spatial-temporal treatment of emission maps, underestimated emissions from anthropogenic activities, the coarse wind data and/or the smoothing of actual extreme values in the simulation.

Line 187-194. Although applying the offset seems to improve the comparison between observations and simulations, the model is not capturing well the observed CH_4 variability (R^2 is too small), thereby model and observations are not reflecting the same air masses, sources (industries or natural processes). The authors superficially mention the possible influence of the tropopause height in the simulations, but without quantifying this impact. Have the authors considered the possible influence of the PBL? Or the shape of the constant a priori profile used for EM27 retrievals? Please include

a more detailed discussion of the possible reasons for these discrepancies.

Response: Thanks for your valuable comments. For this CH_4 bias, we attribute to the errors in the troposphere height and the general offset from CAMS dataset in Sect.3.3. For the vertical distribution of our case, the upper pressure is up to 50 hPa and the simulated RBL height varies from approx. 200 m (in the morning and at night) to 1500 m (during the daytime). And we use the vertical distribution of water vapour to check the simulated PBL height and conclude that the PBLs are situated in the right vertical layers. With the restriction of the relatively coarse vertical layers (26 layers) in our vertical distribution, the PBL heights potentially hold some uncertainty. The CH_4 vertical profile changes with the increase of PBL heights, which leads to the error in XCH_4. The constant a priori profile shapes in EM27/SUN retrievals (while the actual atmospheric profiles are variable) do attribute to the smoothing error of measurement values.

A major offset in modelled CH_4 concentration fields could potentially be attributed to the errors in the troposphere height and a general offset from CAMS. In the CH_4 vertical concentrations profile, we find that the typical sharp decrease is given within the tropopause height. Tukiainen et al. (2016) also find the similar sharp decrease when using the air-cores to retrieve the atmospheric CH_4 profiles in Finland. During the simulation, the background concentration values of CAMS are directly fitted to the WRF pressure axis, without the consideration of the actual tropopause height, thus this is unlikely to be the case. An illustration of the vertical distribution for CH_4 is provided in Appendix C. While the CO_2 vertical distribution shows a quite flat decrease with the increase of pressure and there is no need to consider the tropopause height during the grid treatment in the vertical layer. Then in view of CAMS, the reports from Monitoring Atmospheric Composition and Climate (MACC) described that CAMS held an increase of bias and RMSE (approx. 50 ppb) in each part of the world, compared to the Integrated Carbon Observation System (ICOS) observations in 2017 (Basart et al., 2017). Galkowski et al (2019) also mentioned one CH_4 offset (approx. 30 ppb within troposphere), when initializing the concentration fields using CAMS. Apart from these two major potential reasons for the bias, the influence from the in-accurate simulated planetary boundary layers and the shape of the constant a priori profile used for the retrievals both potentially contribute to the discrepancies for the concentration fields. Due to the lack of fine measured vertical concentration profiles, it is not easy to quantify these errors and attribute these potential reasons to this 2.7% error quantitatively. Thus, a DCM-based analysis is presented in Sect.4, aiming at eliminating the bias from these relatively high initialization values for CH_4 and making it easier to assess WRF-GHG results with respect to the measurements.

Line 195-235: As mentioned in the general comments, the discussion and attribution of GHGs sources is very simple. Please consider to include a robust statistics analysis to support the main conclusions of this section.

Thanks for your suggestions. In this study, our focus is on the model for the major emission sources. We would like to know the major contributions for urban areas and test whether the DCM can work well to highlight the influence from human activities and weaken the impact from the strong biogenic emissions. Thus, the main conclusions for the tracer analysis are demonstrated (Line 370-380) that 'The biogenic component plays a pivotal and leading role in the variations of XCO_2. The impact from anthropogenic emission sources is somewhat weak compared to this, while for XCH_4, the enhancement is dominated by human activities.' The WRF-GHG mesoscale modelling framework is being built in Munich based on the first worldwide permanent column measurement network of Munich. The quantification analysis for different emission tracers is suggested to run for this Munich in which more emission tracers (e.g., biogenic emissions from wetland for XCH_4, the traffic emission and strong point sources' emissions in urban areas, parts of the anthropogenic emissions) are being added and more data are available. In this Munich case, it is possible to get the long-period data and we can do the 'real' simulation for anthropoogenic emission, instead of the simulated values entirely based on the priori data.

Line 248. Have the authors analyzed the vertical distribution of the winds within PBL for the comparisons?

The vertical layers in WRF model are defined following the pressure definition and the top-layer pressure (50 hPa) is already up to tropopause. PBL normally varies from a few hundred meters (morning and night) to over one thousand meters (daytime, pick at noon) which are situated from the second vertical layer (morning and night) to 13^{th} layer (noon) of our domain, corresponding (see the left column of Fig.3 below). The wind is higher than the surface wind. The WRF-GHG outputs do provide these simulated PBL values but there is lack of observations to assess these simulated values. Avolio et al., (2017) did some sensitivity analysis of PBL in the WRF model with a case study of southern Italy, and concluded that the simulated PBL heights are mostly overestimated.

Line 260-289. A plot showing the CH4 and CO2 enhancement observed and simulated as a function of the wind

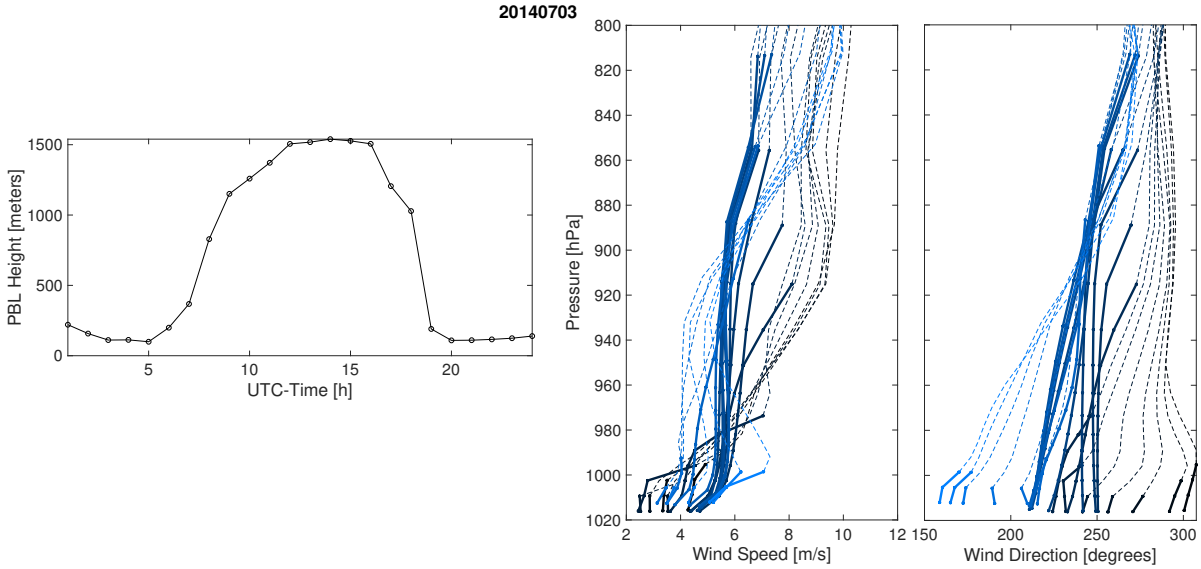

Figure 3: Variations of the simulated Planetary Boundary Layer (PBL) Height (left side), and the wind speeds (middle) and wind directions (right) within PBL on July 3rd July in Tegel. The colors from black to blue represent the time from morning to evening. The bold solid lines represent the values within PBL.

directions or differences between simulated and observed wind directions could help to explain better the results of the section 4. Regarding to Figure 7 and 8, why are not the wind fields considered in Figure 7 similarly to Figure 8? Why does not Figure 8 include the 1st July? Why are not the performance values for CO_2 included in the text?

Thanks for your suggestions. We did several tests to find whether the biases of ΔXCO_2 and ΔXCH_4 between measurements and simulations are associated with wind fields or differences between measured and simulated wind directions. A dependency on wind speeds or wind directions is not found in this study (like the scatter-plot (Fig.9) provided by Vogel et al., (2019)). But in our case, we find that the extreme comparison points in the right column of Fig.8 always correspond to higher discrepancies of wind directions between measurements and simulations. These contents are included in Sect.4.1 and the Fig4 and Fig.5 of this response (Fig.7 and Fig.8 of the content):

To further understand the differences of ΔXCO_2 and ΔXCH_4 between measurements and simulations (see Fig.7, right column and Fig.8, left), the comparison of hourly mean ΔXCO_2 and ΔXCH_4 values for these four targeted dates is illustrated in the right column of Fig.8. Due to the restriction of measured wind information, we illustrate the difference of simulated and measured wind directions at 10 meters (i.e. Fig.2(b)) with respect to the hourly mean ΔXCO_2 and ΔXCH_4. We find that the real hourly mean ΔXCO_2 and ΔXCH_4 values are generally higher than the simulated values. Extreme points are colored by red and blue in the right column of Fig.8, standing for large differences between measured and simulated wind directions at 10 meters. We see that a large difference of wind directions is a necessary but insufficient condition for the bias of ΔXCO_2 and ΔXCH_4 between measurements and simulations. In future studies, this may be verified further.

Then to make the consistence with targeted simulation dates for each comparison case, we target 3rd, 4th, 6th and 10th July in this case (see Fig.7, Fig.8, Fig.9 and Fig.10, Sect.3.3 and Sect4).

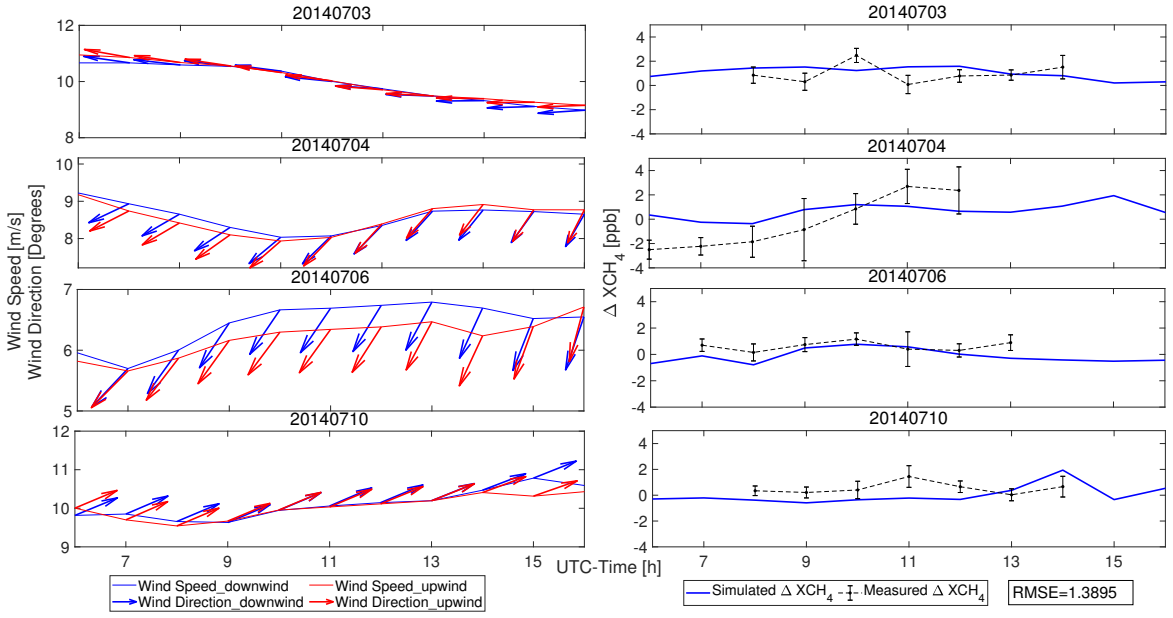

Figure 4: Modeled wind fields for downwind (blue lines) and upwind (red) sites (left column), and downwind-minus-upwind differential evaluation for measured (blue) and simulated (black) XCH_4 (right column) on 3rd, 4th, 6th and 10th July 2014. Based on the selection of downwind and upwind sites in Table.1, ΔXCH_4 is calculated using Eq.6, 7 and 8, depicted by blue lines for measurements and black lines for simulations. The black error bars in the right column are the standard derivations of the minute values of the hourly mean.

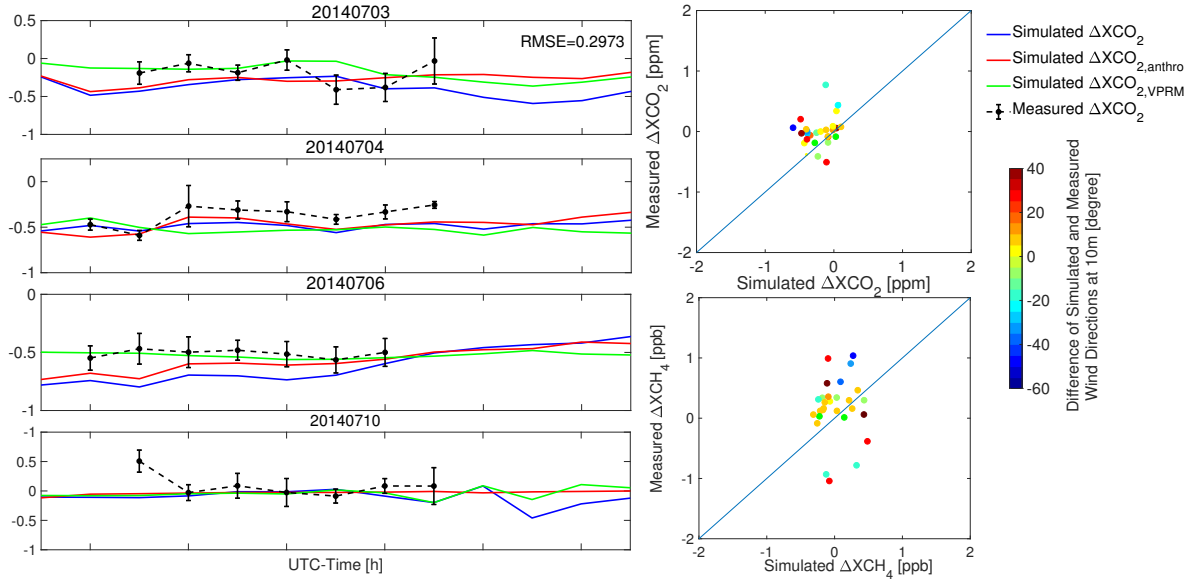

Figure 5: Measured (black lines) and simulated (blue) ΔXCO_2 on 3rd, 4th, 6th and 10th July 2014, and Comparison of hourly mean ΔXCO_2 and ΔXCH_4 for these four days. The ΔXCO_2, calculated using Eq.5, 6 and 7, are depicted by blue lines in the right column. The red and green lines show the variation of the differences between downwind and upwind sites in XCO_2 changes from anthropogenic and biogenic activities, respectively. The points color in right column are coded by the difference of the simulated and measured wind directions at 10 meters.

References

Avolio, E., Federico, S., Miglietta, M. M., Feudo, T. L., Calidonna, C. R., Sempreviva, A. M.: Sensitivity analysis of WRF model PBL schemes in simulating boundary-layer variables in southern Italy: an experimental campaign. Atmospheric research, 192, 58-71, https://doi.org/10.1016/j.atmosres.2017.04.003, 2017.

Beck, V., Koch, T., Kretschmer, R., Marshall, J.and Ahmadov, R., Gerbig, C., Pillai, D., and Heimann, M.: The WRF Greenhouse Gas Model (WRF-GHG). Technical Report No. 25, Max Planck Institute for Biogeochemistry, Jena, Germany., 2011.

Deutscher, N. M., Notholt, J., Messerschmidt, J., Weinzierl, C., Warneke, T., Petri, C., Grupe, P., and Katrynski, K.: TCCON data from Bialystok (PL), Release GGG2014R1, TCCON data archive, hosted by CaltechDATA, https://doi.org/10.14291/tccon.ggg2014.bialystok01.R1/1183984, https://tccondata.org, 2014.

DuVivier, A. K., Cassano, J. J.: Evaluation of WRF model resolution on simulated mesoscale winds and surface fluxes near Greenland. Monthly Weather Review, 141(3), 941-963, https://doi.org/10.1175/MWR-D-12-00091.1, 2013.

Gierczak, T., Talukdar, R. K., Herndon, S. C., Vaghjiani, G. L., Ravishankara, A. R.: Rate coefficients for the reactions of hydroxyl radicals with methane and deuterated methanes, The Journal of Physical Chemistry A, 101(17), 3125-3134, https://doi.org/10.1021/jp963892r, 1997.

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

Hase, F., Frey, M., Blumenstock, T., Groß, J., Kiel, M., Kohlhepp, R., Mengistu Tsidu, G., Schäfer, K., Sha, M., and Orphal, J.: Application of portable FTIR spectrometers for detecting greenhouse gas emissions of the major city Berlin, Atmospheric Measurement Techniques, 8, 3059–3068, https://doi.org/10.5194/amt-8-3059-2015, 2015.

Jee, J. B., Kim, S.: Sensitivity study on high-resolution WRF precipitation forecast for a heavy rainfall event. Atmosphere, 8(6), 96, https://doi.org/10.3390/atmos8060096, 2017.

Notholt, J., Petri, C., Warneke, T., Deutscher, N. M., Buschmann, M., Weinzierl, C., Macatangay, R., and Grupe, P.: TCCON data from Bremen (DE), Release GGG2014R0, TCCON data archive, hosted by CaltechDATA, https://doi.org/10.14291/tccon.ggg2014.bremen01.R0/1149275, https://tccondata.org, 2014.

Pillai, D., Buchwitz, M., Gerbig, C., Koch, T., Reuter, M., Bovensmann, H., Marshall, J., and Burrows, J. P.: Tracking city CO_2 emissions from space using a high-resolution inverse modeling approach: a case study for Berlin, Germany, Atmospheric Chemistry and Physics, 16, 9591–9610, https://doi.org/10.5194/acp-16-9591-2016, 2016.

Ridgwell, A. J., Stewart J. M. Keith G.,: Consumption of atmospheric methane by soils: A process-based model. Global Biogeochemical Cycles, 13.1, 59-70, https://doi.org/10.1029/1998GB900004, 1998.

Vogel, F. R., Frey, M., Staufer, J., Hase, F., Broquet, G., Xueref-Remy, I., Chevallier, F., Ciais, P., Sha, M. K., Chelin, P., Jeseck, P., Janssen, C., Té, Y., Groß, J., Blumenstock, T., Tu, Q., and Orphal, J.: XCO_2 in an emission hot-spot region: the COCCON Paris campaign 2015, Atmos. Chem. Phys., 19, 3271-3285, https://doi.org/10.5194/acp-19-3271-2019, 2019.