We are grateful to the reviewers additional items that have been requested. The new elements require the addition of a new coauthor, S.Riette.

Anonymous Referee #3

Overview: The manuscript presents simulated carbon dioxide fields for one week centered over Paris. The work demonstrates and tests the ability of a high-resolution meso-scale model to reproduce observed meteorological and carbon dioxide dynamics, with a focus on urban areas, Paris in particular. This work is appropriately placed in ACP, and contributes to the burgeoning area or studying carbon emissions from urban areas. I have some general and specific concerns delineated below, after satisfactorily addressing these issues I would recommend publication.

General Comments: Overall things look quite nice and interesting, but I have a couple of larger reservations that require more work and must be addressed.

1) CO2 boundary condition. This is only briefly touched upon in section 2, is unclear, and seems inadequate. From what I understand the model is initialized with a flat field based on observations, and then run from there. Is there any spin-up time? What is done with air flowing into the domain (what value is it assigned)? What impact do varying boundary condition choices make on simulations? We know that in regional studies boundary conditions play a tremendously important role (Lauvaux et al. TELLUS 2012). The authors must better described what they've done for boundary conditions, and make quantitative assessments of impacts of boundary condition choices on simulations.

Authors : In terms of CO_2, the first day of the period is initialized at 00UTC with a flat field based on observations at EIF : at 00UTC, the measurement value at the Eiffel Tower, above the BL, can be considered as a background value. The other days, the predicted CO_2 field from the end of the previous day is used for initialization.

The spin-up time concerning the meteorological fields is very short (less than 2 hours) because initial fields come from AROME, at 2.5km resolution, possessing its own data assimilation. The mesoscale portion of the kinetic energy spectrum already exists in the initial fields and therefore develops rapidly in Meso-NH. Concerning the spin-up of CO_2, the first day, we can suppose that, for anticyclonic conditions and weak winds, nocturnal CO_2 in the BL is mainly driven by anthropogenic and biogenic emissions, then by horizontal transport (but small with weak winds) and lastly by the vertical transport (negligible during the night). Considering an initial homogeneous vertical profile of CO_2 on 21 March 00 UTC would only lead to misrepresent, at the growing stage of the BL, the entrainment at the top of the BL of CO_2 that could have been trapped in the residual layer at the end of the previous day, but this is a hypothetical situation. The spin-up of CO_2 the other days is negligible, as we use the Meso-NH forecast of the previous day.

At the lateral boundaries, we apply a constant value homogeneously on the vertical given by the CO_2 background concentration measurement at Eiffel Tower (minimum value of the day), slightly different every day.
A sensitivity test on lateral boundaries for CO_2 has been leaded by using CO_2 fields from LMDZ model (with a horizontal resolution 0.83° ×1.25° (latitude × longitude) over Europe), like in Ahmadov et al. (2009). But the impact is negligible on CO_2 prediction fields over the Paris region.

We agree that in the future, this aspect could be improved by restoring the lateral boundaries toward a large-scale analysis of CO2, like MACC analysis. Also, it is thought that for such anticyclonic situations, the effect of large-scale advection of CO2 in the boundary layer are probably weak compared to the vertical turbulent diffusion. Additionally, the simulation domain is sufficiently large to minimize the lateral boundary condition effects over Paris area, as the domain includes the main pollution sources influencing air quality over Paris region, like Lille, industrial area of Benelux and port activities of the Normandy coast.

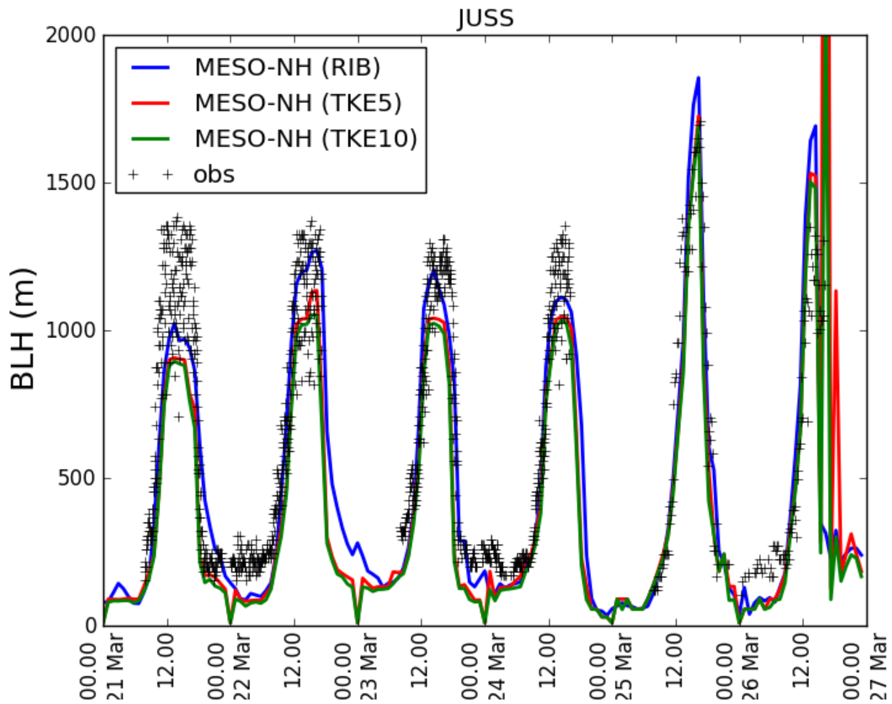
On continue to the paragraph P 28161 line 24 :
"The first day's CO_2 field was initialised with the CO2 background concentration measurement at Eiffel Tower (minimum value of the day), with a homogeneous vertical profile, horizontally consistent across the entire model domain, while the other days used the predicted CO2 field from the end of the previous day as a starting concentration field. The boundary conditions CO2 profiles during each day's simulations were also taken from the homogeneous vertical profiles."

We propose to add P 28162 line 1 :
"A sensitivity test on lateral boundaries for CO2 has been leaded by using CO2 fields from LMDZ model (with a horizontal resolution 0.83° ×1.25° (latitude × longitude) over Europe), like in Ahmadov et al. (2009) but shows no significative impact on CO2 prediction fields over the Paris region. In the future, this aspect could be improved by restoring the lateral boundaries toward a large-scale analysis of CO2, like MACC analysis. Also, it is thought that for such anticyclonic situations, the effect of large-scale advection of CO2 in the boundary layer are probably weak compared to the vertical turbulent diffusion. Additionally, the simulation domain is sufficiently large to minimize the lateral boundary condition effects in the domain of interest, as it includes the main pollution sources influencing air quality over Paris region, like Lille, industrial areas of Benelux and port activities of the Normandy coast."

Referee : *2) Overall, there are qualitative statements describing how 'small' errors are and how well model represents things, but there is little to not quantitative substantiation. Errors of 100m in nocturnal pbl height are not small by any measure. This is 50% or more of the observed pbl height. Qualitative statements should be toned down.*

Authors : We completely agree and we have added quantitative statements and revised the redaction.
The diagnosis of BLH has been revised as the TKE method based on 10% of the near surface value (noted TKE10 on Fig.1) does not seem optimal. It has been compared to the same method with 5% of the near surface value (noted TKE5), and to the bulk Richardson number approach (Seibert et al., 2000), considering a critical value of 0.25 (Sorensen et al., 1997) (noted RIB). The latter gives the best estimates, as can be seen on the following figure (that will not be added to the manuscript) at Jussieu for the REF simulation. The advantage of the RIB method is confirmed for all the stations and all the scores (biases, rmse, R^2) and is adopted for the revised version of the paper. Table 2 has been amended (see below), including the correlation coefficient R^2 .
Figure 6 has been modified according to that, and the subfigure with the sensible heat fluxes has been removed as the measurements were specific of an area of grassland vegetation and not representative of the general area (and therefore not comparable to the mean predicted sensible heat flux of the 2km grid mesh).

Also, in response to your question 4., a more complete quantitative evaluation of the BLH over Trappes (SW of Paris) has been added by comparison to radiosoundings over one year. It gives a bias of 19m at 12h UTC (12h forecast) and –6m at 00h UTC (24h forecast). Therefore, a new figure (Fig.7) and a new Table (Table 3) have been added.



<u>Figure (Not included in the revised version) :</u> Time series of BLH at JUSSIEU with the 3 methods for the REF simulation compared to the observation.

The proposed modifications in the text are :

L7 P 28162 : The phrase is added :

" In order to generalize the evaluation of the BLH, the REF simulation has been daily run for one year (August 2010 – July 2011) over the same domain, in exactly the same configuration"

L21-23 P 28166 : Replacement of :

" The diagnosis of the BLH in the model is based on the TKE profile (the first level from the ground with a TKE less than 10% of the near surface value determines the BLH) (Seibert et al.,2000)."

By :

"The diagnosis of the BLH in the model is based on the bulk Richardson number approach (Seibert et al., 2000), considering a critical value of 0.25 (Sorensen et al., 1997)".

Sørensen J.H., Rasmussen A. and Svensmark H., 1997a: Forecast of Atmospheric Boundary Layer Height Utilised for ETEX Real-time Dispersion Modelling. Physics and Chemistry of the Earth.

L15-27 P 28167 : Replacement of :

« The REF simulation captures reasonably well the BLH for all the sites during daytime, with negative biases between 85m and 122m (Table 2). In the morning, the onset time of the ABL mixing and the growth rate of the BLH are particularly well reproduced. Maxima of BLH are also well captured, except a slight underprediction the first two days for the 3 sites. The increase of daytime BLH on 25 March is correct on the urban and sub-urban sites, while it is overestimated on the rural site, as well as on 26 March for the 3 sites.

During nighttime, the REF simulation represents fairly well the shallow mixing depth over urban and sub-urban sites, with only small negative biases of 47~m and 34~m respectively (Table 2). Underestimations occur at JUSS on 25-26 March, and at SIRTA on 21, 24-25 and 25-26 March (but BLH measurements at SIRTA are not so reliable on 25-26 March).

The small underprediction of the nocturnal BLH for these 3 nights at SIRTA site is also visible on the sensible heat flux that is slightly underestimated (Fig.6.d)."

By :

"The REF simulation captures reasonably well the BLH for all the sites with correlation coefficients between 0.89 at JUSS and 0.71 at TRN (Table 2). During daytime, biases are negative, between 8m (at TRN) and 70m (at JUSS). This can be explained for the rural site by the small negative bias on T2M (Fig.3). In the morning, the onset time of the ABL mixing and the growth rate of the BLH are particularly well reproduced (Fig.6). Maxima of BLH are also well captured, except a large underprediction the 1st day for the 3 sites (up to 300m at JUSS) and a small one the 4th day at JUSS and TRN. The increase of daytime BLH on 25 March, compared to the other days, is predicted at the 3 sites, but slightly underestimated at SIRTA and overestimated at TRN.

During nighttime, the REF simulation represents fairly well the shallow mixing depth over urban and sub-urban sites, but tends to underestimate it slightly (negative biases of 45m and 5m respectively, Table 2).

The evaluation of BLH has been generalized over the one year period by comparisons against BLH from daily soundings at TRAP, also estimated with the same critical bulk Richardson number. Correlation are presented in Fig.7 with the regression line included, and biases and rmse are reported in Table 3. Statistics reveal a very good agreement at this sub-urban site, with biases of +19m and –5m for 12H (12UTC soundings) and 24H (00UTC soundings) forecasts respectively. But we can underline that statistics on SIRTA and TRAP agree on the fact that the model tends to underestimate slightly the nocturnal BLH at sub-urban site. The mean diurnal cycle exhibits a good agreement between observation and REF at noon and midnight (Fig.7.c). "

JUSSIEU				
	MEAN day		**MEAN night**	
OBS	867		222	
	BIAS day	**RMSE day**	**BIAS night**	**RMSE night**
REF	-70	222	-45	131
R^2	0.89			
RUR	-226	320	-117	132
R^2	0.78			
SIRTA				
	MEAN day		**MEAN night**	
OBS	731		155	
	BIAS day	**RMSE day**	**BIAS night**	**RMSE night**
REF	-34	256	-5	127
R^2	0.76			
RUR	-160	313	-44	106
R^2	0.68			
TRAINOU				
	MEAN day			
OBS	661			
	BIAS day	**RMSE day**		
REF	-8	303		
R^2	0.71			
RUR	-23	293		
R^2	0.72			

Table 2 : Statistical scores of the BLH for the REF and RUR simulations, compared to the observation. Night-time is considered from 19UTC to 8UTC.

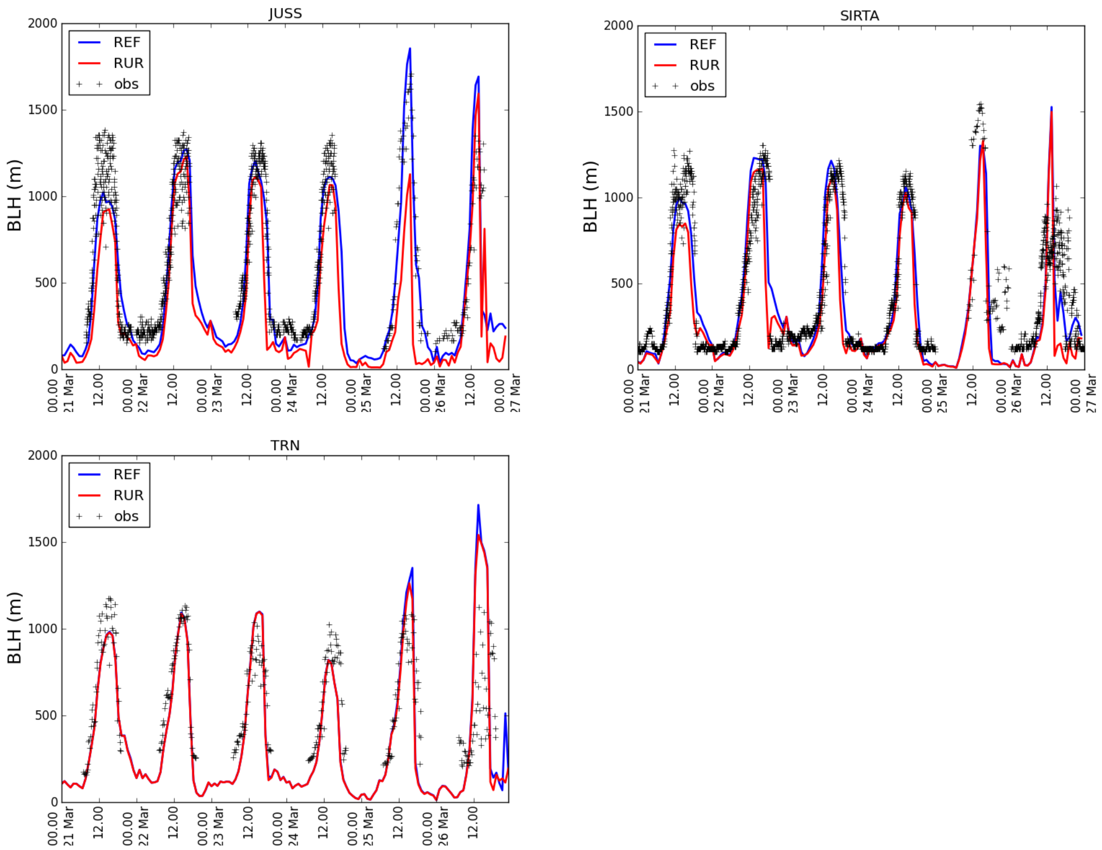
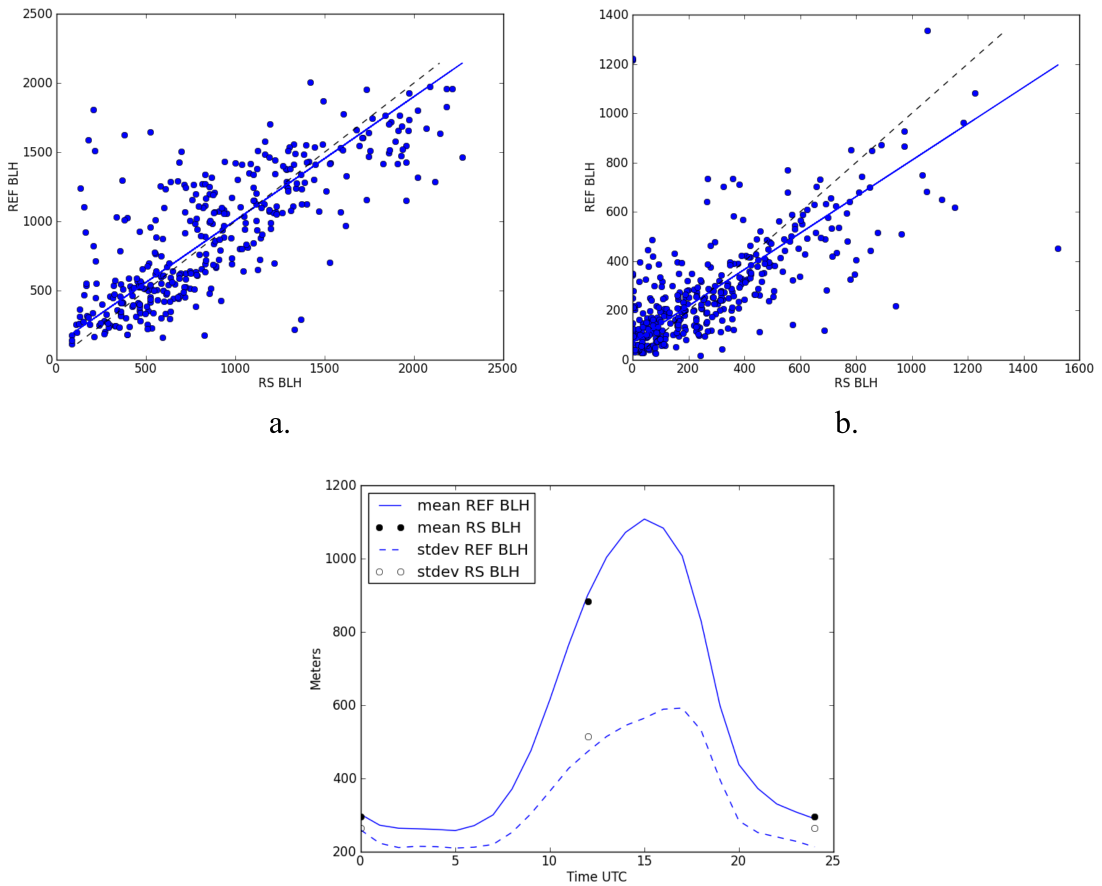


Figure 6 : Time series of BLH (in meters above ground level (AGL)) for March 21 - 26 at JUSS (a, urban site), SIRTA (b, sub-urban site) and TRN (c, rural site).



c.

Figure 7 : a. and b. : Correlation between observed (noted RS) and simulated BLH at Trappes for the REF simulation for one year (August 2010-July 2011) for 12H forecast (12 UTC sounding) and 24H forecast (00UTC sounding) . The regression is indicated by the continuous line. C : Diurnal cycle over the year of the BLH (continuous line for the Mean and dashed line for the standard deviation) at Trappes predicted by REF with the observed values marked by dots.

TRAPPES (BLH in m)				
	MEAN 12UTC	**STD DEV 12 UTC**	**MEAN 00UTC**	**STD DEV 00UTC**
OBS	883	515	296	265
REF	902	473	290	213
	BIAS 12H forecast	**RMSE 12H forecast**	**BIAS 24H forecast**	**RMSE 24H forecast**
REF	+19	337	-6	191

Table 3 : Statistical scores of the BLH at Trappes from observation (mean and standard deviation, noted STD DEV) and from the REF simulation over 1 year (August 2010-July 2011).

Also, L17-25 P 28168, due to the modified results of BLH with the Bulk Richardson number method, replacement of :
"The comparison between REF and RUR simulations on the BLH (Fig.6) shows that both predict similar daytime BLH on urban and sub-urban sites except for 25 March, as it is largely underpredicted over JUSS and SIRTA without TEB. Therefore the biases of BLH for the RUR simulation during daytime are twice the ones of the REF simulation on JUSS and SIRTA, with a rmse also increased (Table 2). But systematically, the RUR simulation underpredicts the nighttime BLH on the urban and sub-urban sites (doubled biases), showing the effectiveness of the TEB scheme in representing the storage of heat in urban materials during the night. The impact at the sub-urban site is smaller but not negligible all the nights."

by :

"The comparison between REF and RUR simulations on the BLH (Fig.6) shows a systematic reduction of the BLH at the urban site during the day and during the night, degrading significantly the negative biases and the correlation. At the suburban site, the difference between both is reduced compared to the urban site but not negligible, as evidenced by the statistics, especially on the maximum of the afternoon underestimated. At the rural sites, curves are combined. This comparison demonstrates the effectiveness of the TEB scheme in representing urban-rural contrasts on the BLH."

Referee : *Also, statements attributing all mismatch to surface flux misrepresentation need to be restructured.*

Authors : Yes, we agree. L25-27 P28167 has been removed :
"The small underprediction of the nocturnal BLH for these 3 nights at SIRTA site is also visible on the sensible heat flux that is slightly underestimated (Fig.6.d)."

Referee : *Nowhere has it been shown that remaining transport errors do not explain discrepancies, at least in part. In fact, nowhere has the impact of different transport errors on CO2 fields been shown. This would be a very valuable exercise to do, demonstrating quantitatively the impact of some of the associated transport errors.*

Authors : We agree that this test would be very informative. But to our knowledge, evaluating the impact of transport errors is possible with a lagrangian scheme but not with an eulerian model. We do not see the possibility to lead this kind of impact study with Meso-NH. However, different representations of resolved and subgrid transports can be tested in an eulerian model. Concerning the resolved transport, we have compared the upwind WENO (3^{rd} order) scheme to the centred advection schemes (4^{th} order on the wind). In addition to the reduced computation time[1], the uncentred schemes give also the best scores on the meteorological variables (T2M, HU2M and 10m winds). Because we wanted to keep the best available meteorology in our simulations, we did not evaluate further the CO2 concentrations produced by the 2 different sets of numerical schemes and have not compared them statistically. Concerning the subgrid transport, the 1D turbulence with the Bougeault-Lacarrere (1989) turbulent length seems optimal at 2km resolution. Nevertheless, the impact of transport errors has been better underlined in the part 5 and in the conclusion, as it will be shown.

[1] The physics is called with a time step 10 times longer, and the advection with a time step 2 times longer

Referee : *3) There needs to be more/better presentation of the model and observed CO2 values (such as in 1:1 plots). As currently presented, it is difficult to assess model performance. Conclusions about 'small' errors attributable to transport need to be toned down. Errors of 10+ ppm at night are not small, and even errors of a couple ppm in daytime could be quite significant in an inverse modeling sense.*

Authors : Yes, we completely agree that assessments were inadequate, as well as some comments. We have added :

- statistics for T2M
- R^2 values for BLH
- an annual evaluation of BLH at Trappes with 1:1 plots
- 1:1 plots and statistical scores (bias, rmse, R^2) for CO2.

Part 5. has been rewritten, as well as a part of the abstract and the conclusion, as it is presented after.

Referee : *4) This study focuses only on one week of modeling and observations. Conclusions thus must be quite limited, as one cannot extrapolate to generalized model performance from such a limited duration comparison, which could be particularly favourable or unfavourable. The limited duration of model/observations must be presented, and its impact on conclusions should be discussed. One element of this is discussing time/computation to simulate one-week, and whether the current model construct could be expected to run for years to compare w/ the observational record being recorded in Paris & Europe.*

Authors : You agree with this point. Therefore, and with an objective *in fine* of inverse modelling for the CO2-MAGAPARIS project, the model has been run in exactly the same REF configuration (domain, resolutions, initialization and coupling, duration) over one year, from August 2011 to the end of July 2012. In terms of time computation, one day of simulations costs 6h CPU on the NEC supercomputer of Meteo-France, and therefore one year represents around 2200h CPU, which is not too much expensive. These good performances have been made possible with the Runge-Kutta time splitting of the temporal scheme associated to uncentred advection schemes (PPM and WENO), allowing a reasonable time step (see Table 1.). The one year run offers an evaluation of the BLH by comparisons with soundings at Trappes (See 2.), whereas an evaluation of CO2 over the year has not been possible for the moment.
We agree that the limited duration comparison between CO2 observation and modelling does not allow to generalize the model performance but the daily runs over one year, that have been added in the revised version, provide a very good evaluation of the BLH in sub-urban area that gives confidence in the modelling database.

The **conclusion** was changed from line 24 Page 28174 to the end :

Instead of : "The small discrepancies are mainly linked to weak errors on the vertical transport for the ground stations located in the Paris plume (e.g. GIF), or on the horizontal transport for ground stations in the plume of an airport (e.g.GON), and also on the too coarse resolution of the anthropogenic inventories. The performance of the urban parameterisation scheme TEB is crucial to reproduce the UHI, the urban-rural contrasts and the nocturnal BLH on urban and sub-urban sites, and consequently the mixing ratio maxima.
The good representation of CO2 mixing ratio on urban and sub-urban sites during nighttime emphasizes the use of the modelling system in inverse framework with nocturnal surface and tower station records.

The study also demonstrates the potential of the CO2-MEGAPARIS stations to be used for inverse methods, as the stations are devoted to monitor long term measurements of CO2 and offer an adequate and comprehensive database to quantify surface fluxes. The next step of the study is now to apply one year of Meso-NH forward modelling in the same configuration with the CO2-MEGAPARIS measurement network for inverse methods."

We propose :
"Discrepancies on nocturnal CO2 concentrations are consecutive to vertical transport errors, with a mean negative bias of 5m on the BLH over SIRTA during nighttime, and also maybe to horizontal transport errors and to the spatio-temporal inaccuracy of anthropogenic emissions over the city and the airports. The urban parameterisation scheme TEB proved crucial to reproduce the UHI, the urban-rural contrasts and the CO2 diurnal cycle.
The limited duration comparison between CO2 observation and modelling does not allow to generalize the model performance but the daily runs over one year provide a very good evaluation of the BLH in sub-urban area that gives confidence in the modelling database. This one year of Meso-NH forward modelling can now be used for inverse methods based on the CO2-MEGAPARIS measurement network. The study also demonstrates the potential of the CO2-MEGAPARIS stations to be used for inverse methods, as the stations offer an adequate and comprehensive database to quantify surface fluxes and are devoted to monitor long term measurements of CO2. However, additional CO2 stations, especially ground stations in denser parts of the city, would be beneficial."

Referee : *Specific Comments*:

Referee : *Title: add 'the' as "Modelling over the Paris Region.."*
Authors : OK

Referee : *Abstract: rephrase opening line, accurate simulation is very useful, but not necessarily 'essential' (some data driven methods may answer many of the relevant questions without simulations)* :
Authors : "Accurate simulation of the spatial and temporal variability of tracer mixing ratios over urban areas is challenging, and interesting in order to utilize CO2 measurements in an atmospheric inverse framework to better estimate regional CO2 fluxes."

Referee : *Abstract: restate nocturnal BLH only slightly underestimated. Errors of 100m may be 50% or more at night, not accurate to present this as 'slightly underestimated'*
Referee : *Abstract: 'mainly linked to the misrepresentation' this should be rephrased, these biases are 'likely' linked to errors in anthropogenic sources, but you have not definitively shown that in the work here.*
Referee : *Abstract: sentence starting 'The CO2 cycle at these sites. . ." what is the impact of the pbl bias on observations? Should be mentioned here*

Authors : The abstract has been corrected from line 14 to the end :

Instead of :
"Boundary layer heights (BLH) at urban, sub-urban and rural sites are well captured, especially the onset time of the BLH increase and its growth rate in the morning, that are essential for tall tower CO2 observatories. Only nocturnal BLH at sub-urban sites are slightly underestimated a few nights, with an error less than 100m. At Eiffel tower, the observed spikes of CO2 maxima occur every morning exactly at the time at which the atmospheric

boundary layer (ABL) growth reaches the measurement height. The timing of the CO2 cycle is well captured by the model, with only small biases on CO2 concentrations, likely linked to the misrepresentation of anthropogenic emissions, as the Eiffel site is at the heart of traffic emission sources.
At sub-urban ground stations, CO2 measurements exhibit maxima at the beginning and at the end of each night, when the ABL is fully contracted, with a very strong spatio-temporal variability. The CO2 cycle at these sites is generally well reproduced by the model, even if some biases on the nocturnal maxima appear in the Paris plume partly due to small errors on the vertical transport, or in the vicinity of airports due to small errors on the horizontal transport (wind direction). A sensitivity test without urban parameterisation removes UHI and underpredicts nighttime BLH over urban and sub-urban sites, leading to large overestimation of nocturnal CO2 concentration at the sub-urban sites. The agreement of daytime and nighttime BLH and CO2 predictions of the reference simulation over Paris agglomeration demonstrates the potential of using the meso-scale system on urban and sub-urban area in the context of inverse modelling."

We propose :
"Boundary layer heights (BLH) have been evaluated on urban, sub-urban and rural sites during the campaign, and also on a sub-urban site over one year. The diurnal cycles of the BLH are well captured, especially the onset time of the BLH increase and its growth rate in the morning, that are essential for tall tower CO2 observatories. The main discrepancy is a small negative bias over urban and sub-urban sites during nighttime (respectively –45m and –5m), leading to a few overestimations of nocturnal CO2 concentrations at sub-urban sites. The diurnal CO2 cycle is generally well captured for all the sites. At Eiffel tower, the observed spikes of CO2 maxima occur every morning exactly at the time at which the atmospheric boundary layer (ABL) growth reaches the measurement height. At sub-urban ground stations, CO2 measurements exhibit maxima at the beginning and at the end of each night, when the ABL is fully contracted, with a strong spatio-temporal variability.
A sensitivity test without urban parameterisation removes UHI and underpredicts nighttime BLH over urban and sub-urban sites, leading to large overestimation of nocturnal CO2 concentration at the sub-urban sites. The agreement between observation and prediction for BLH and CO2 concentrations and gradients, both day and night, demonstrates the potential of using the urban meso-scale system in the context of inverse modelling."

Referee : *Intro: sentence starting 'Indeed, with 12 millions of inhabitants" change millions to 'million', and word 'largest' between third and megacity, and rephrase 'Moscow), and is estimated to emit about 14%...'*
Authors : : OK

Referee : *Intro: modify to 'Moreover, it is an ideal test location. . .'*
Authors : OK

Referee : *Intro, Paragraph 2: Would be appropriate here to acknowledge urban CO2 studies being pursued with different methods on different cities (Indianapolis: Gurney KR et al., Environ. Sci. Technol. 2012; Salt Lake City: Strong C et al., JGR 2011; Los Angeles: Kort EA et al., GRL 2012)*
Authors : Page 28157 line 27 on continue :

"Urban CO_2 studies have been recently pursued with different methods on different cities (Indianapolis: Gurney et al., 2012; Salt Lake City: Strong et al., 2011; Los Angeles: Kort et al., 2012)".

Referee : *Section 2, Paragraph 1: The concluding sentence is confusing and incorrect 'we will improperly call mixing ratio by concentration. This should be removed and should be corrected throughout later on* :
Authors : OK : "concentration" has been replaced by "mixing ratio" everywhere.

Referee : *Section 2 Paragraph 2: typo: 'miximum' should be 'maximum'*:
Authors : OK

Referee : *Section 2 final paragraph: Why replace urban with rock? What is the reason for choosing rock? This should be explained.*

Authors : The rock doesn't induce evapotranspiration that would modify the dynamics of the BL, but it presents a significant roughness and non erodible elements, like vegetation, on the contrary to the sand, for instance.

P 28162 lines 4-7 : Replacement of :

"(2) the second simulation (RUR hereafter) is conducted without the TEB urban scheme (the urban land-use covers are replaced by rock, treated by ISBA) in order to quantify the effect of the urban parameterisation ; "

By :

"(2) the second simulation (RUR hereafter) is conducted without the TEB urban scheme, in order to quantify the effect of the urban parameterisation : the urban land-use covers are replaced by rock, treated by ISBA, as the rock doesn't induce evapotranspiration that would modify the dynamics of the BL, and presents a significant roughness and non erodible elements, like vegetation."

Referee : *Section 3: There needs to be more detailed explanation of observations. I would like to see more detailed explanation of sampling. Are observations being made wet or dry? What is the calibration strategy used? What scale are observations placed on? What are estimated accuracies/precisions/biases?*

Authors : P.28163 line 5 : on continue

In the 3 CO_2-Megaparis stations, the observations were made wet and a correction on water vapour was applied using the dedicated Picarro analyzer software. All observations were calibrated against the NOAA X2007 scale. Each station was equipped with a calibration and target gas tanks unit owning specific peculiarities. Concerning the GIF and TRN stations, as part of the ICOS infrastructure, an automated gas chromatographic system (HP-6890) was operated for CO_2 measurements of ambient air (Gibert et al., 2007). A detailed explanation on the calibration strategy and accuracy/precision estimates is under preparation in an article from Xueref-Remy et al.(2013). The precision for the different datasets is given in Table 1. The temporal sampling is 1h for GIF and TRN stations, and 5min for EIF, GON and MON.

	EIF	GON	MON	GIF	TRN
Precision	0.382ppm	0.065 ppm	0.101 ppm	0.5 ppm	0.5 ppm

Table 1. Precision of the dataset.

Gibert, F., M. Schmidt, J. Cuesta, P. Ciais, M. Ramonet, I. Xueref, E.Larmanonou, and P. H. Flamant (2007) : Retrieval of average CO2 fluxes by combining in situ CO2 measurements and backscatter lidar information. J. Geophys. Res., 112, D10301.

Referee : *Section 4.1 final statement: this is qualitative. How does this matter for CO2? It seems quite large, not small. Also, in the figure, the model rural-semiurban exhibit greater difference than observations. This isn't discussed at all, and I am curious as to why this is, and how this type of error may impact CO2 fields.*

Authors : Following your comment, a quantitative assessment has been introduced.

	URBAN		
	BIAS	**RMSE**	**R^2**
REF	+ 0.8	1.6	0.9
RUR	-3.0	2.9	0.75
	SUB-URBAN		
	BIAS	**RMSE**	**R^2**
REF	0.0	1.0	0.96
RUR	-0.5	1.4	0.92
	RURAL		
	BIAS	**RMSE**	**R^2**
REF	-0.2	1.5	0.93
RUR	-0.2	1.5	0.93

Table 2 : Evaluation of the mean Bias, Rmse (in °C) and correlation coefficient R^2 between observations and REF and RUR simulations on T2M at Montsouris (URBAN), SIRTA (SUB-URBAN) and Trainou (RURAL) sites

You are right that there are small differences on the rural-suburban contrasts between REF and observations (OBS), but they are occasional, and despite this they are well reproduced most of the time. It is confirmed by statistics as biases between REF and OBS on T2M are 0° and –0.2° and rmse are 1° and 1.5° for semi-urban and rural stations respectively (Tab.2). T2M is directly linked to the sensible heat fluxes that drive the BL development, and therefore the CO2 mixing. In terms of CO2, it will only influence biogenic emission but not anthropogenic emission. Therefore, only TRN could see an impact of surface temperature error on CO2 emissions. But the too small sampling at TRN (only the 1st day available with 1h temporal sampling) doesn't allow to establish statistics.

The modifications are :

P 28165 lines 26-28 :

"The REF simulation reproduces well the increasing trend of the temperature and the urban-rural contrasts (Tab.2, with correlation R^2 between 0.9 and 0.96 for the 3 stations), with only a systematic overestimation of the maximum temperature at the urban site of 2°C (inducing a mean bias of +0.8°C)."

P 28168 lines 11-16 :
"On fig.4.c, the RUR simulation underestimates systematically the urban temperature (Tab.2 with a negative bias of –3°C, and the corrections by the analysis at 00UTC are important) and removes the UHI : the small differences between the three sites are only linked to the orography effect of the Paris basin and to the cooling associated to the evapotranspiration for the rural site compared to the rock replacing the urban area in the RUR simulation."

Referee : *Section 4.3 typo, fix 'bassin' to 'basin'*:
Authors : OK

Referee : *Section 5.1 'its representation at local scale could be improved with finer emission inventories' This is likely true, but has not been demonstrated, and the role of transport has not been quantified or eliminated. Need to tone down this (and other similar statements) to acknowledge transport may still play a role.*

Authors : We agree and Section 5.1 has been rewritten as presented below.

Referee : *5.2: 'these small errors are not attributed to the vertical transport. . .' as in comment above, this statement extends beyond findings presented here, as transport has not been eliminated as a significant player.*
Authors : The new version of 5.2 (presented below) takes into account your remarks.

Referee : *5.2 paragraph 4: typo, should be 'is not negligible (Fig. . .'*
Authors : OK

Referee : *5.3 'between 1 and 2 ppm' Where do these numbers come from? This seems exceedingly important and is not at all evident from Figure 11. This discussion and substantiation should be expanded.*
Authors : Fig.9 and Fig.11 have been improved in terms of visibility, and the redaction has been reformulated (see below).

Referee : *Figure 1: Should reduce scale extent on topography. Also, all figures should have units labeled with the scale.*
Authors : OK

Referee : *Figure 2b: The model seems far to variable and low compared to observations that are more uniform. This, and the impact on CO2 simulations, should be discussed.*

Authors : The error is on the order of values of the statistical bias calculated for the 235 stations, with a mean deficit of 1°C at 11UTC. It is also in the order of values of errors of operational Numerical Weather Prediction models. The impact of an error of 1°C in rural areas on the CO2 is not straightforward to estimate, as photosynthesis depends mainly on surface temperature and humidity.

Referee : *Figure 3 (and Fig. 7): There is a large deviation in bias errors in the morning hours with sunrise. This is hardly discussed in the text. Potential reasons for this and its import should be mentioned. Furthermore, this seems to maybe point to errors in boundary layer growth, which in the text is characterized as being very well represented.*
Authors : The link between errors on T2M and BLH is not so obvious : during the day, the bias on T2M is negative for rural station and could explain the small negative bias of the BLH on TRN (-5m) : it has been mentioned above. But for urban and sub-urban stations, the bias is positive for T2M and negative for BLH so the link is not straightforward. During the night, the bias is negative for urban and suburban BLH whereas it is slightly positive for T2M, so again the link is not direct.

Referee : *Figure 4: These plots need to be all properly lined up and of higher quality.*
Authors : This has been corrected (see below).

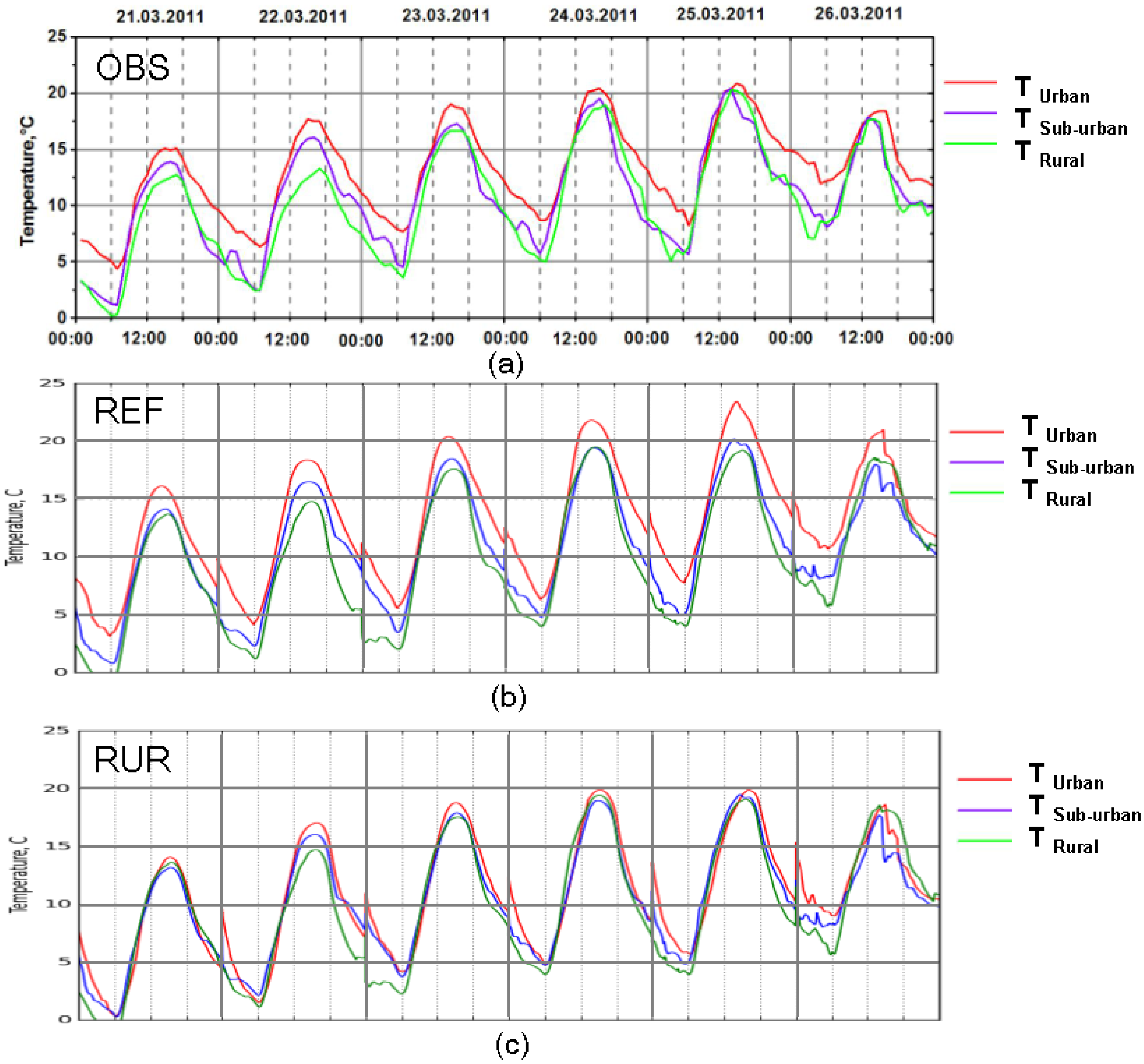


Figure 4. Diurnal variation of hourly near surface air temperature (in °C) at urban, suburban and rural stations measured (from Pal et al., 2012) (a), predicted by REF simulations (b) and RUR simulations (c)

Referee : *Figure 6: There are a number of features discrepancies here not addressed in the text. The timing on simulated vs observed BLH does not seem that great (for instance at TRN*

), and may indeed even be erroneous at JUSS (see the 23rd). REF & RUR simulated pbl heights are extremely similar most of the time, in contrast to the text statement. Further confusing is they appear to produce the same sensible heat flux at SIRTA, and both seem in significant error (far too high sensible heat), but the pbl height looks reasonable. This discrepancy needs to be explored & explained further.
Authors : All this part has been corrected with the modified diagnostics of BLH based on the Bulk Richardson numbed. Figure 6 has been modified (see above) as well as its description, and the comparison on the sensible heat fluxes has been removed as explained above.

Referee : *Figure 8 & 10: I would like to see maybe an average daily cycle of CO2 as well. A 1:1 plot of model and observations would be very informative as well.*
Authors : Your suggestions have been very useful as these 2 kinds of plots have been introduced and have improved the comprehension/validation.

Referee : *Time series are good but prevent more quantitative assessments. Scales of Fig 10 b and d are far too expanded to assess model performance.*
Authors : You are completely right. Therefore a new figure with 1:1 plots for CO2 mixing ratios (presented below as New Fig.11) has been included. And the Fig.10 has been replaced by the mean daily cycle for the 5 stations.
Part 5.1 and 5.2 have been rewritten as presented below.

New part 5 :

CO_2 mixing ratio predictions are investigated herein using time series of predictions from REF, RUR and also NAN simulations against observations, for the Eiffel Tower (hereafter EIF), Gonesse (GON), Montge-en-Goelle (MON), Gif-sur-Yvette (GIF) and Trainou (TRN). The NAN simulation allows to distinguish the sites quasi-fully influenced by anthropogenic emissions (EIF) and those strongly influenced by anthropogenic emissions (GON and GIF), to the site both exposed to anthropogenic and biogenic emissions (MON) and finally to the rural site quasi fully driven by assimilation and plant transpiration (TRN). It is worth noting that peak values of anthropogenic emissions over Paris and its airports occur during rush hours, between 5 and 8UTC (local wintertime), and 18 and 22UTC (not shown). But nocturnal emissions remain important at CDG Airport as it is today the leading European airport traffic at night, with an average of 170 movements per night (almost 15% of the total of the airport traffic over 24 hours).

5.1 Evaluation at the urban site : Eiffel Tower at 300m of height

The mean diurnal cycle is presented in Fig.8.a. for EIF. The observed CO_2 maxima occur much later than for the other sites, generally between 09 and 11UTC. While the other sites record the highest concentrations when the BLH is fully contracted, the Eiffel Tower concentrations show maxima during the late morning as the ABL expands. As JUSS is close to EIF, observed and predicted BLH evolutions at JUSS are used to help analyzing CO_2 observations and predictions at EIF (Fig.9). The observed CO_2 spikes trigger exactly at the time (vertical dashed line) at which the growing BLH reaches the measurement height of the Eiffel Tower (310 m as shown in Fig.9.a). These spikes have a very short duration as the ABL grows quickly, favoring the rapid mixing of pollutant in a deeper layer and consequently the rapid CO_2 mixing ratio decrease.
In terms of timing and temporal evolution, the modelled mixing ratios can be seen to agree well with observations : predicted and observed maxima occur at the same time, meaning that the predicted BLH reaches 310m at the right time. The predicted CO_2 peaks are also very brief, in agreement with measurements. The correct timing is confirmed on the mean diurnal cycle (Fig.8.a).
However, a few discrepancies appear in the temporal evolution. Firstly, the longest period of high observed mixing ratios has occurred during the night of 22-23 March. It is underestimated, by the model, probably due to an underprediction of the BLH (measurements at JUSS were not available), even if the REF simulation tends to produce higher BLH than for the other nights, reaching punctually the EIF measurement height. Secondly, another discrepancy occurs on 25 March at 2UTC, as the model predicts a peak of 450ppm that does not occur in reality, associated to a reservoir of pollutant in the simulated residual layer. All of this can explain that statistically, the comparison model *vs*. observation gives a negative bias of 6 ppm (rmse of 17 ppm) and a middling 0.35 correlation coefficient (Fig.10.a).
In terms of intensity, there are small biases on CO_2 mixing ratios, that could be partly linked to the misrepresentation of the anthropogenic emissions and to horizontal transport errors. This is illustrated on the strongest peak event measured at EIF during the campaign (25 March at 11UTC) as a consequence of the negligible wind during all the night and the early morning (Fig.11). Anthropogenic CO_2 accumulates over Paris intra-muros in the Seine valley in the shallow early morning ABL (Fig.11.c at 8UTC) and this reservoir reaches 300m height with the ABL growing at 11UTC (Fig.11.d). The model underpredicts the maximum over Eiffel Tower but reproduces CO_2 mixing ratio magnitudes at 300m comparable to the measurements magnitude over the eastern part of Paris town (Fig.11.f with measurement in

coloured square). The predicted plume mixing ratios are directly linked to the CO2 emissions that are higher on the eastern part of Paris. It is therefore likely that the underestimation at EIF is partly due to the too coarse anthropogenic emissions, as the correct mixing ratios have been produced on another part of Paris, and partly to horizontal transport errors, frequent with weak winds. Moreover, the simulated ground level mixing ratios closely match the lower observed mixing ratio values at the sub-urban sites (Fig.11.e). So the general anthropogenic pollutant accumulation over Paris city on 25 March is correctly reproduced, and its representation at local scale could probably be improved with finer emission inventories.

The RUR tends to delay of 30 minutes the peak of CO2 (Fig.8.a), as the growing phase of the BLH is delayed by the same time (Fig.8.a). The misrepresentation of nocturnal UBL with RUR does not impact CO2 concentration at EIF, as the measurement is located above. Overall the statistics are worse, with a correlation coefficient of 0.05.

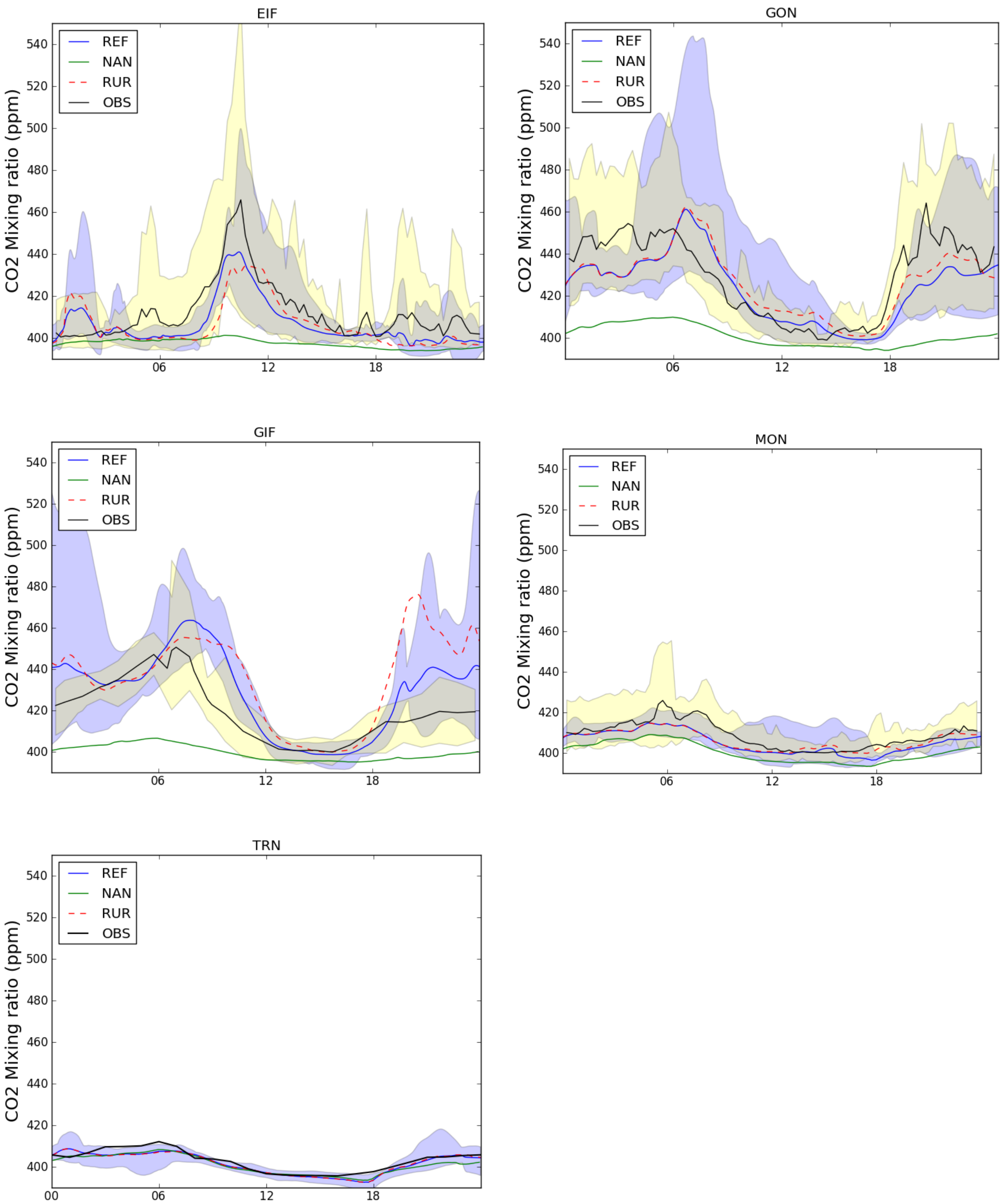


Figure 8 : Mean temporal evolution (over 6 days) of the CO2 mixing ratios (in ppm) measured (black line) and predicted with REF (blue line), RUR (red line) and NAN (green line) simulations at EIF (a), GON (b), GIF (c), MON (d) and TRN (e). The yellow area is between the minimum and maximum of the measurements, and the blue area between the minimum and maximum of the REF simulation.

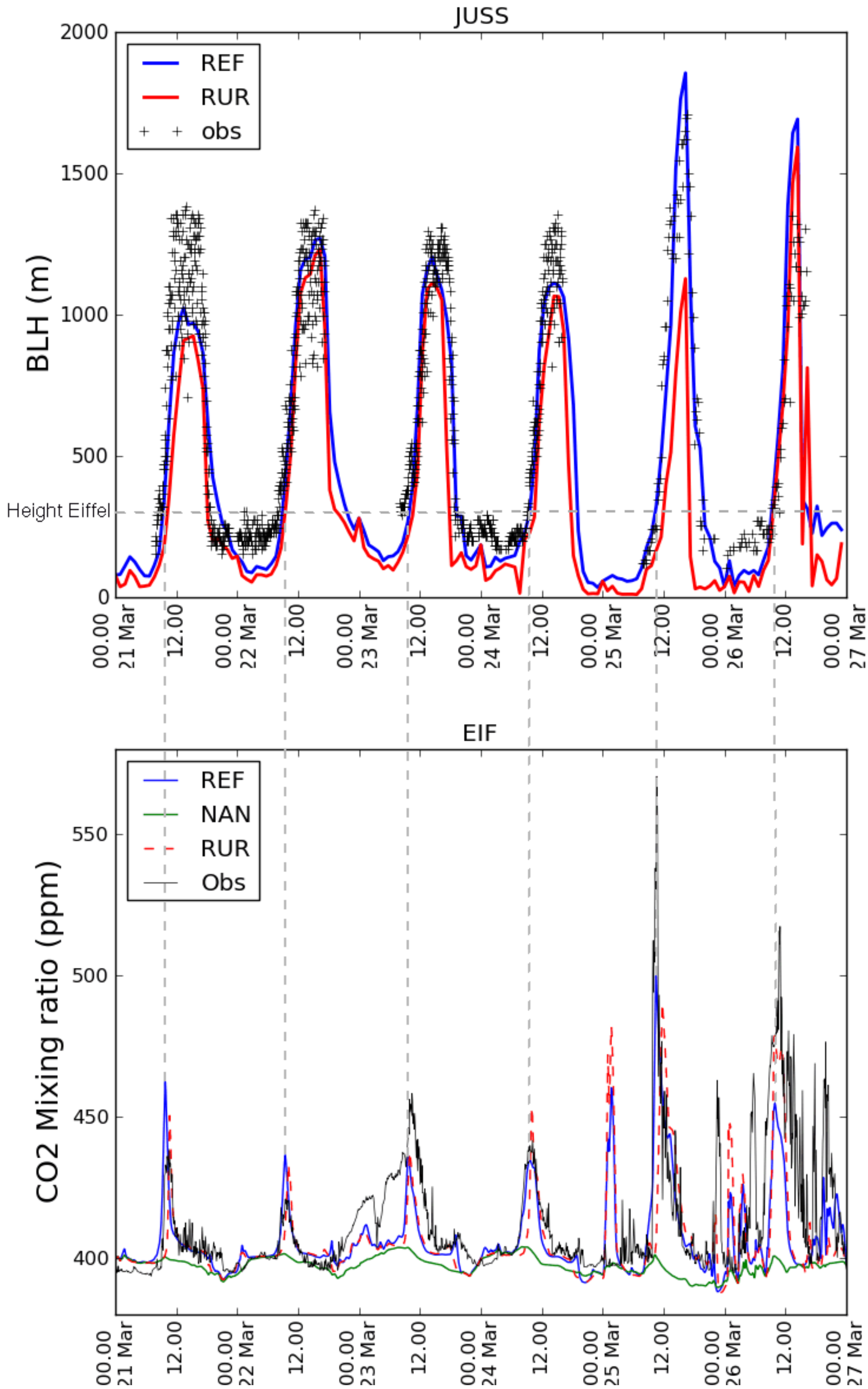


Figure 9 : (a) Time series of BLH predictions at Eiffel and observations at JUSS (in meters above ground level, AGL) for REF (blue) and RUR (red) simulations. (b) Time series of CO_2 predictions and observations (in ppm) at EIF for REF (blue), RUR (red) and NAN (green) simulations. The vertical dashed lines correspond to the time in the morning at which observed BLHs reach 310m.

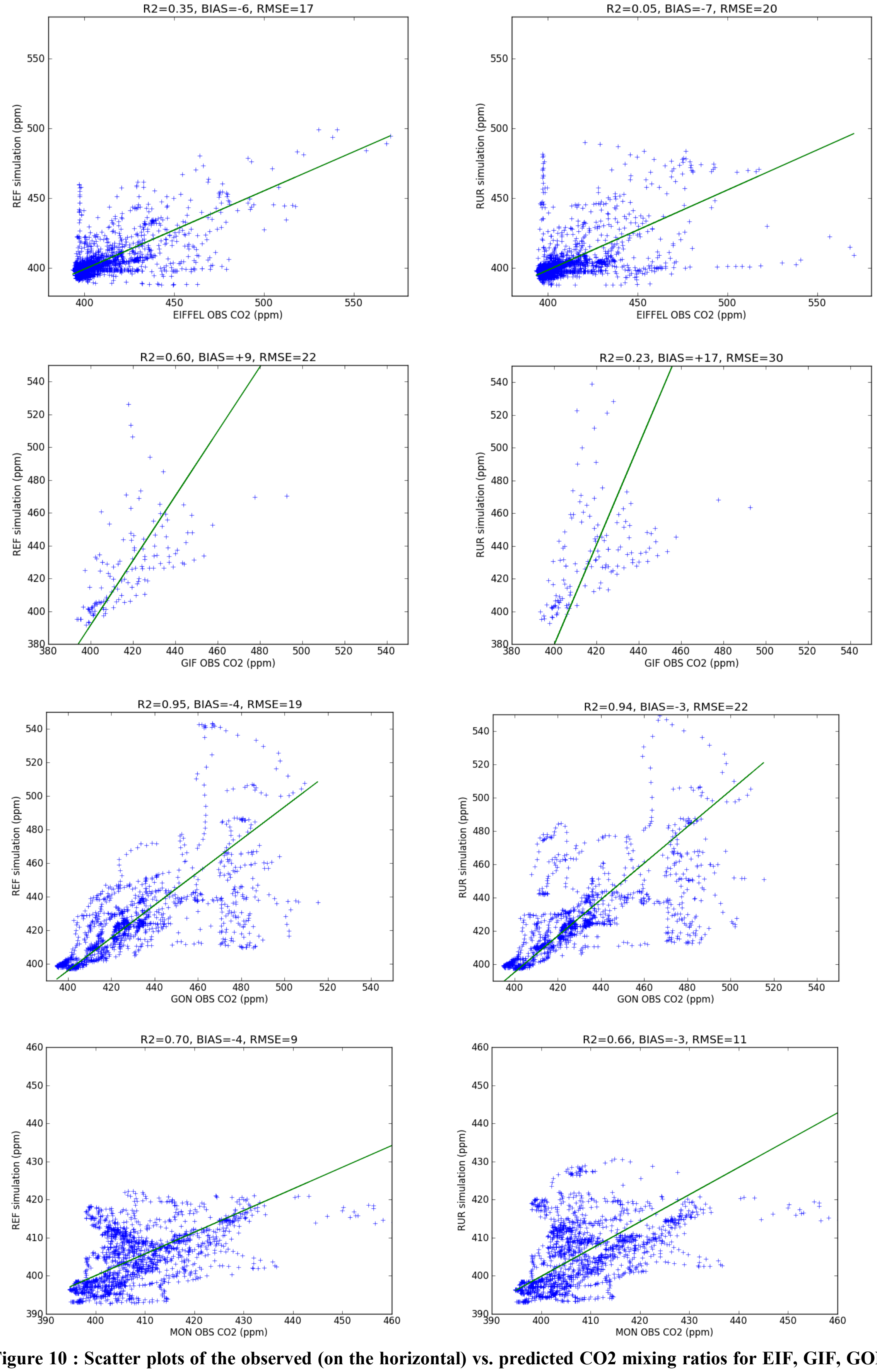


Figure 10 : Scatter plots of the observed (on the horizontal) vs. predicted CO2 mixing ratios for EIF, GIF, GON and MON for REF simulation (on the left) and RUR simulation (on the right), with an orthogonal linear regression

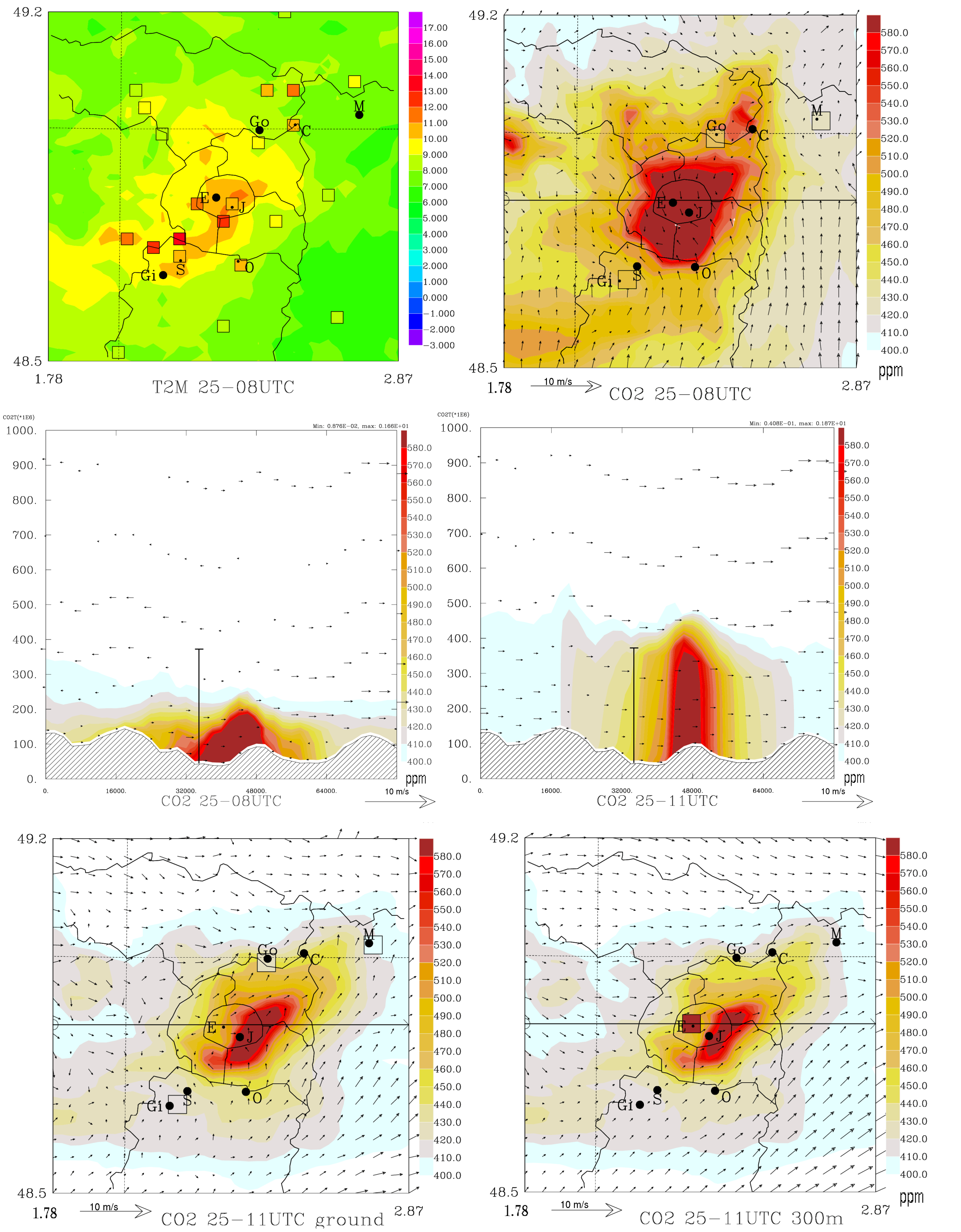
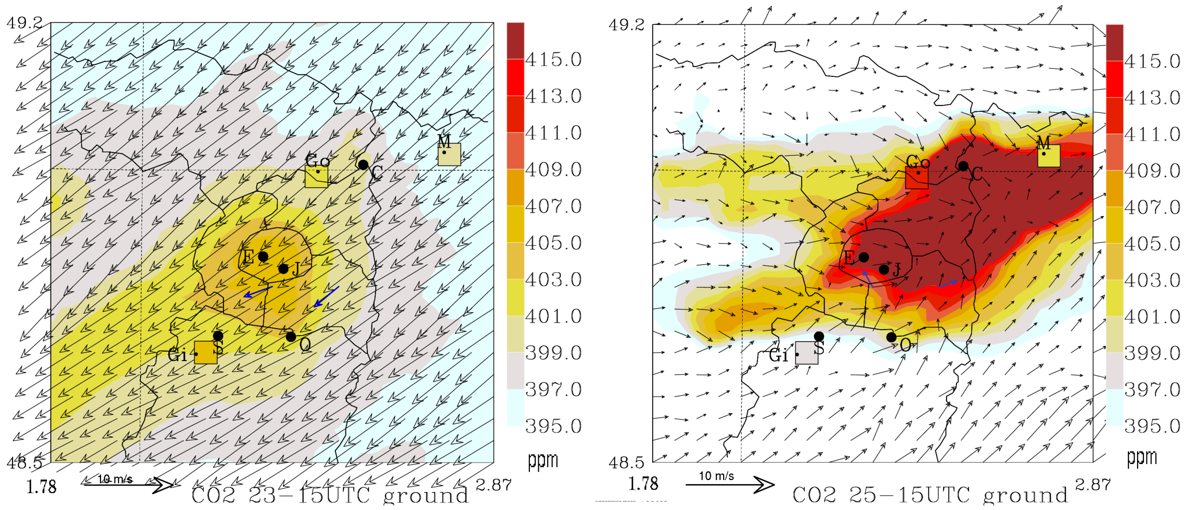


Figure 11 : MESO-NH predictions : For March 25 at 08UTC (a) 2 m temperature (in °C), (b) Horizontal cross section of CO2 mixing ratio (in ppm) near the ground. (c) Vertical cross section of CO2 mixing ratio (in ppm) according to the axis given in (b) for March 25 at 08UTC, and for March 25 at 11UTC (d), with wind vectors superimposed. The Eiffel tower is symbolized by a stick, with a length corresponding to its measurement height. Horizontal ticks indicate meters. For March 25 at 11UTC : Horizontal cross-section of CO2 mixing ratio (in ppm) near the ground (e) and at 300~m height (f) with wind arrows superimposed. Coloured squares correspond to observed CO2 mixing ratio.

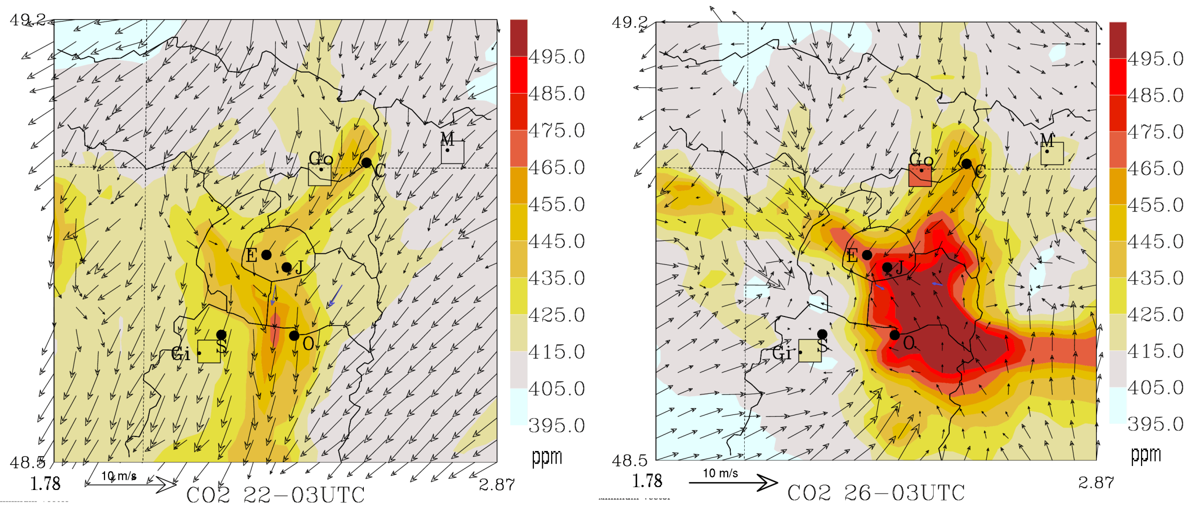
5.2 Evaluation at the sub-urban and rural sites

CO_2 mixing ratio observations and predictions at the sub-urban and rural sites are presented in Fig.8 in terms of diurnal cycle and compared for statistics in Fig.10.
On the contrary to EIF altitude station, the surface sub-urban sites (GON and GIF) always measure maxima in the second part of the night and in the early morning, when the BLH is strongly contracted. They exhibit a strong temporal variability of CO_2 mixing ratio (yellow area in Fig.8).
At GIF site (Fig.8.c), the REF simulation reproduces correctly the timing of the diurnal cycle of CO_2 mixing ratio. But if the minimal CO_2 mixing ratios are well captured by the model, the nocturnal maxima tend to be overestimated, inducing a positive bias of 9 ppm, and a correlation coefficient of 0.6 (Fig.10.c). This can be directly linked to the vertical transport error, as SIRTA exhibited a negative bias of 5m on the BLH.
The RUR simulation degrades significantly the statistics (R^2 is equal to 0.23 and bias to +17) as shown on the diurnal cycle : the reduced mixing in the BL without TEB extends the period of strong CO_2 at the morning and at the end of the afternoon, and the lower nocturnal BLH increases the concentrations.
At GON, observation and REF simulation are in fairly good agreement with a correlation of 0.95 and a small negative bias of 4 ppm (Fig.10.e), and the diurnal cycle is well reproduced (Fig.8.a). The discrepancies mainly concern the maximum of the CO_2 peak in the early morning and the temporal evolution (not shown) insures that only 25 March morning is imputed. On 25 March at 08 UTC, the near ground temperature on the North-East of Paris (Fig.11.a) is underestimated, inducing an error on the vertical transport leading to an overestimation of the mixing ratio (Fig.11.b). On the contrary, REF tends to underestimate the nocturnal concentrations (Fig.10.b). During a major part of the March period, GON has undergone the plume of CDG airport during the night, in an east flux, as the airport keeps a night traffic activity. Therefore the horizontal transport on one side, and the uncertainties on the emission on the other side, are two potential sources of error of CO_2 at this station.
The MON station is classified as a rural site, but is nevertheless influenced by anthropogenic emissions from Paris and CDG airport, as the difference between REF and NAN simulations is not negligible (Fig.8.d). The period exhibits two regimes, with a quite regular diurnal cycle the first 4 days and north-east winds that protect the site from Paris and CDG plumes, as on Fig.12.a. and Fig.13.a., and a stronger variability the last two days due to the weak winds with variable directions including mainly westerly winds (Fig.12.b. and Fig.13.b). The model reproduces fairly well the CO_2 concentration, with a correlation of 0.7 and a negative bias of 4 ppm (Fig.10.g), but the second period was more exposed to horizontal transport errors and emission uncertainties. This is underlined by the statistics on the RUR simulation, that unusually tends to improve the scores (Fig.10.h), meaning that vertical transport errors are less involved.
While almost no observations are available for the rural site of TRN during this period, the measurements at the beginning of the period allow to check the predicted mixing ratio. The CO_2 diurnal cycle is almost identical each day of the period, with a nocturnal maximum due to the ecosystem respiration (Fig.8.e), and a CO_2 mixing ratio decrease in the ABL when the BLH increases, due to CO_2 vertical mixing but also to photosynthesis activity which depletes the boundary layer CO_2 mixing ratio. The three simulations REF, RUR and NAN are almost superimposed, meaning that the vegetation fully drives the diurnal cycle of carbon.



a. b.

Figure 12 : Horizontal cross-sections of CO2 mixing ratio (in ppm) and predicted wind arrows (in m/s) near the ground for March 23 at 15UTC (a.), 25 March at 15UTC (b).



a. b.

Figure 13 : Horizontal cross-sections of CO2 mixing ratio (in ppm) and predicted wind arrows (in m/s) near the ground for 22 March at 03UTC (a) and 26 March at 03 UTC (b) of BLH (in m above ground level) (on the left) and CO2 (on the right). Observations of CO2 are added in coloured squares, and wind in blue arrows. White colours correspond to values less than the minimum coloured one.

5.3 CO2 horizontal heterogeneity in the afternoon and the night for inversion purposes

Figure 8 shows that the model reproduces well the midday lower mixing ratios at the different sites. Even if strong convective mixing in the ABL during daytime induces lower mixing ratio values, urban-rural contrasts can lead to significant horizontal gradients and also moderate variability from one day to another. For instance, on 25 March at 15UTC, observed horizontal CO2 gradients reach up to 15ppm between GIF and GON, and this is quite well reproduced by the model. The predicted mixing ratio over MON is overpredicted by 10ppm, but the station is located on the border of the predicted plume. Both the insufficient spatial accuracy of the anthropogenic emissions and errors on the horizontal transport (observed winds are in blue arrows) could explain it. The situation differs from the previous days when the well-established north-easterly winds dilute the pollutant, smoothing the CO2 gradients and inducing the maximum mixing ratio values of the measurement stations at GIF site (Fig.12.a). Errors on the predicted winds are small as well as on CO2. Table 4 presents mean CO2 mixing ratios along a rural-urban transect during the period 13H-17H UTC for the 6 days with non negligible gradients. These gradients are well represented by the model with the urban scheme, and a little less with the RUR simulation. This demonstrates the possibility to apply inversion during daytime on urban and sub-urban area for ground and altitude stations.

Rural-urban contrasts on CO2 are stronger during the night, as shown in Table 4 for the period 00H-06H. Note that unfortunately EIF should not be considered here as it is located above the UBL. Therefore, there is no dense urban station available here to compute urban-suburban gradients. The gradients are fairly well reproduced by the model, especially when the flux is well-established like during the first four days (Fig.13.a). The complex wind circulation on 26 March at 03UTC involves a stronger variability on CO2, mostly represented by the model (Fig.13.b). It underlines the possibility to apply inversion also in the nocturnal UBL.

		TRN	GIF	EIF	GON	MON
13H-17H UTC	**OBS**	**396**	**401**	**408**	**403**	**401**
	REF	**396**	**400**	**403**	**403**	**400**
	RUR	**396**	**402**	**405**	**406**	**402**
00H-06H UTC	**OBS**	**408**	**430**	404	**446**	**414**
	REF	**406**	**436**	404	**434**	**411**
	RUR	**406**	**438**	404	**434**	**411**

Table 4 : Mean CO2 mixing ratios (in ppm) for the 2 periods 13H-17H UCT and 00H-06H UTC, observed and predicted for the 5 sites TRN, GIF, EIF, GON and MON.