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 #2

The paper of Lac et al presents the results of the mesoscale forward model Meso-NH and validates the model using meteorological and CO2 measurements in Paris and its surroundings within the CO2-MEGAPARIS project. The model exercise is validated using 6 days in March 2011. Different modelling approaches are used in order to investigate the role of the Urban Heat Island on the diurnal evolution of weather variables and the boundary layer height in urban/suburban/rural sites, and the role of urban emissions on the diurnal evolution of atmospheric CO2 mixing ratio in urban and background sites. The paper presents a valuable modelling approach in order to understand the temporal and spatial variability of weather variables and CO2 mixing ratios in urban areas and provide new insights in the urban carbon cycle. However, there are few things in the paper that need to be addressed before its publication.

General comments:

1. The validation of the model approach with observations is done in a very qualitative way making very difficult for the reader to assess of agreement between the model and observations. Moreover, there is a lack of assessment of global performance of the model and observations. I think that reporting coefficients of determination (R2) or 1:1 plots that compare model vs observations would help to better see the model's performance. The description of the temporal evolution of the model performance is described too qualitatively and sometimes the text is difficult to follow. Furthermore, since one of the goals of the paper is to report urban-suburban-rural transects, I think that it would be good to compare such transitions as seen by observations and by the model.

Authors : You are completely right that a more quantitative approach of the validation was necessary with statistics including R^2 and 1:1 plots. Statistics (bias, rmse, R^2) have been included for T2M on urban, suburban and rural sites, R^2 have been added for BLH, as well as statistics on CO2 and 1:1 plots. Also a urban-suburban-rural transect for CO2 was reported in section 5.3. On the basis of your comments in agreement with those of the other Reviewer, all the Part 5. has been reformulated with modified figures. Abstract and conclusion have been partly corrected. This is presented below.

2. The wording of the text will need a revision. There are some sentences difficult to understand and the text is not properly proof-read.

Authors : We have brought some corrections.

Specific comments:

Referee : *Lines 13-19, page 28159. The sentence will need rewording. Not really clear the points why urban areas are challenging in inversion studies. Objectives of the study end page 28159 – beginning 28160. What about the ability of the model in representing the temperature, the relative humidity and the wind fields in the area of the study? I don't fully understand what objective (2) means.*

<u>Authors :</u> Instead of :

“In this context, urban areas are challenging to represent for CO_2 inversion studies : they add to the variability of the BLH, and Angevine (2003) pointed out the important implications of urban-rural contrasts for air quality. But they also present the advantage of a nocturnal boundary layer (NBL) that is mixed compared to the rural one. If the urban effects are well represented, this can limit the errors of the model generally associated to the stable boundary conditions. The performance of urban surface parameterisations are therefore crucial in simulating urban boundary layer (UBL) (Lemonsu et al 2006, Lee et al 2008)”

we propose :

“In this context, urban areas are challenging to represent for CO_2 inversion studies : they add to the variability of the BLH, and Angevine (2003) pointed out the important implications of urban-rural contrasts for air quality. However, this heterogeneity also presents the advantage of a Nocturnal Boundary Layer (NBL) that is mixed compared to the rural one. If the urban effects are well represented, this can limit the errors of the model generally associated to the stable conditions. The challenge is here to be able to simulate all the urban effects with an appropriate urban model. The performance of urban parameterisations are therefore crucial in simulating Urban Boundary Layer (UBL) (Lemonsu et al 2006, Lee et al 2008).”

The objective (2) has been clarified :
Instead of “(2) to infer the effect of urban-rural contrasts on the observed atmospheric CO_2 field” : “(2) to evaluate the effect of urban-rural contrasts on the atmospheric CO_2 field”.

<u>Referee :</u> *Lines 19-21, page 28161. Which is the temporal resolution of the emissions inventory? Which is the spatial and temporal resolution of the CO_2 fluxes?*

<u>Authors :</u> The temporal resolution of the emissions inventory is one hour and it has been added. The spatial and temporal resolution of the CO_2 fluxes correspond to the spatio-temporal resolution of the model, *i.e.* 2km and 1 min.

P 28161, L19-21 have been modified like this :
“The anthropogenic CO_2 emissions are obtained from an inventory (10 km and 1 hour resolutions) provided by University of Stuttgart (Dolman, 2006). Oceanic CO_2 fluxes are parameterised following Takahashi (1997), at the resolution of the model.”

<u>Referee :</u> *Last paragraph of page 28161: Not clear what it means “the boundary conditions CO_2 profiles during each day's simulations were also taken from homogenous vertical profiles”. Are those vertical profiles observed vertical profiles? Or from the model? Why are they homogenous if they are from the previous day's model results?*

You are right that it was not clear. 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.
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.

This has been corrected like this :
"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 Eiffel Tower measure, considering a homogeneous vertical profiles."

Referee : *Furthermore the sampling network hasn't been introduced before. Might it be clearer introducing the CO2-MEGAPARIS sampling network first and then the modelling framework?*
Authors : You are right, the two parts have been reversed

Referee : *Line 16 page 28162. Substitute "served" by "used"*
Authors : OK

Referee : *Line 20, page 28162. Might it be appropriate to say "the French operational meteorological surface network"? First paragraph page 28163. Which is the precision and accuracy of the CO2 observations? Which is the temporal resolution of observations? Are CO2 mixing ratios referred to the International Scale?*

Authors :

Yes, you are completely right on the correction : "the French operational meteorological surface network"
Concerning CO2 stations, CO2 mixing ratios refer to the International Scale. The description has been completed like this :

P.28163 line 5 : on continue

In the 3 CO2-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 CO2 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.

<u>Referee :</u> *Line 8, page 28163: Substitute "leads" by "leaded"*
<u>Authors :</u> OK

<u>Referee :</u> *Line 20-21, page 28163: Propose the following wording for a better understanding "235 stations reporting hourly data for T2M and HU2m and 114 stations reporting daily wind speed and wind direction".*
<u>Authors :</u> Perfect.

<u>Referee :</u> *Line 1-2, page 28164. What does it mean that the evaluation of meteorological simulations are performed in operational weather prediction centres?*
<u>Authors :</u> To clarify these lines, a few precisions have been brought :
"Evaluating meteorological simulations against T2M, HU2M and 10m wind fields is a very common practice in operational weather prediction centers. The scores against screen-level variables are generally difficult to improve and can be considered as very informative of the quality of the surface and boundary layer simulation."

<u>Referee :</u> *Lines 13-15, page 28164. Substitute "for the set of stations" by "for all stations", "wet during the day BUT very good agreement AT night".. "...23 March between 04:00 and 11:00 UT".*
<u>Authors :</u> Perfect for the 2 first corrections. For the third one, it is *"...23 March at 04:00 and 11:00 UT".*

<u>Referee :</u> *Lines 23-25, page 28164. There is no evidence that "the excessive cooling and moistening during the day is mainly attributed to the ISBA scheme".*
<u>Authors :</u> You are right that the formulation is too directive. But considering the direct link between the heat fluxes and T2M/HU2M on one hand, and the differences on scores between urban and rural stations on the other hand, we can consider that *"the excessive cooling and moistening during the day is probably relative to the ISBA scheme".*

<u>Referee :</u> *Lines 3-15, page 28165. The description of the temporal evolution of observed temperature for the campaign must be done using past tenses.*
<u>Authors :</u> You are right.

<u>Referee :</u> *Line 17, page 28165. Suggestion of wording: "Also, the dry conditions during previous days reduced the soil..."*
<u>Authors :</u> Perfect.

<u>Referee :</u> *Lines 25-26, page 28165. Don't understand what it means "the occasional measurements at SIRTA and TRN sites, not taken into account in the operational analysis".*

<u>Authors :</u> It means that measurements at SIRTA and TRN do not include the operational meteorological surface observation network, so data are not taken into account in the data assimilation system to produce the operational analysis. This has been clarified like this :

"It is noteworthy that only the measurements at Paris-Montouris are included in the data assimilation in AROME, inducing the same T2M between observation (Fig.4.a) and initial conditions of the run (Fig.4.b) at 00 UT, on the contrary to the measurements at SIRTA and

TRN sites, that do not include the operational meteorological surface observation network, so are not taken into account in the data assimilation system to produce the operational analysis."

Referee : *Lines 3-4, page 28166. A comparison of R2 between REF and RUR simulations will help the reader to see that REF captures well the urban-rural contrasts. Figure 4c is not described in the text and any word regarding the RUR modelling scheme is made. From the graph is observed that the RUR scheme doesn't capture the urban-rural transect.*
Why not quantify this transect as seen by "obs", "REF" and "RUR"?
Authors :
A table of statistics is introduced to help the reader :

	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

Fig.4.c was commented, but in the part 4.3 on the Importance of the urban scheme : P 28168 Lines 11-16.

Following your comments, 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 : *Line 16, page 28166. Substitute "pointed out" by "pointing out"*
Authors : OK

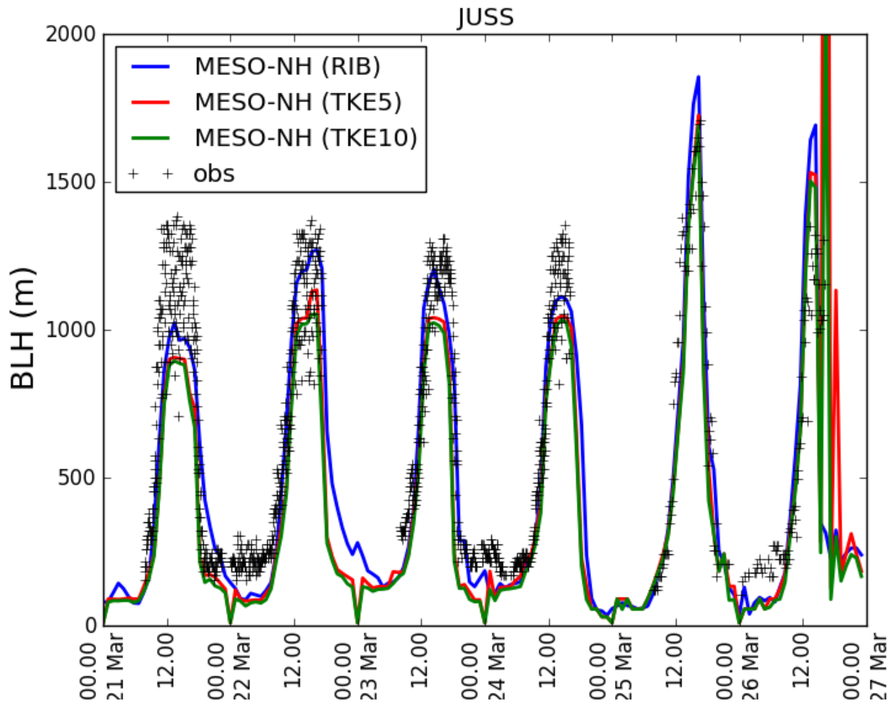
Referee : *Line 23. Page 28166. Substitute "not useful" by "not working"*

Authors : OK

Referee : *Line 25, page 28166. Substitute "was not able to result" by "was not operational"*
Authors : OK

Referee : *Line 4, page 28167. "... mixing LAYER for JUSS, FOLLOWED BY SIRTA and THEN TRN". How different is the boundary layer height for these three sites? The description of the observation should be done in past tenses.*
Authors : Thank you for the correction and the past tense. The difference of the BLH between the sites is given in Table 2 as a mean and commented P 28167 lines 4-14.

Referee : *Lines 17-25, page 28168. REF simulations don't capture the BLH at JUSS for the first 3 days of the campaigns. A sentence about this is needed in the text. Maybe calculating R2 for the entire period would help seeing an increase of the accuracy of the REF simulations in comparison with RUR in terms of reproducing the BLH at urban sites?*
Authors :
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 a question from the other anonymous Reviewer, 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,2200)."

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.

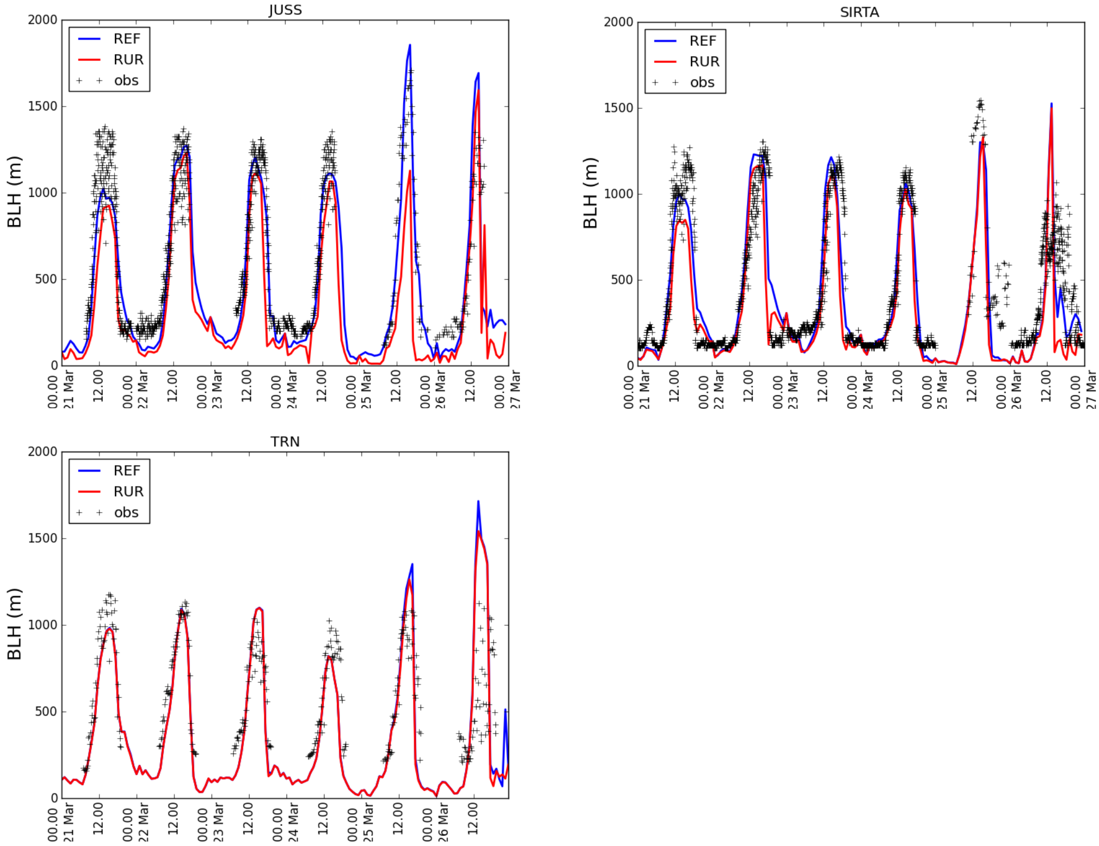
During nighttime, the REF simulation represents fairly well the shallow mixing depth over urban and sub-urban sites, with only small negative biases of 47~m and 34~m respectively (Table 2). Underestimations occur at JUSS on 25-26 March, and at SIRTA on 21, 24-25 and 25-26 March (but BLH measurements 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 :

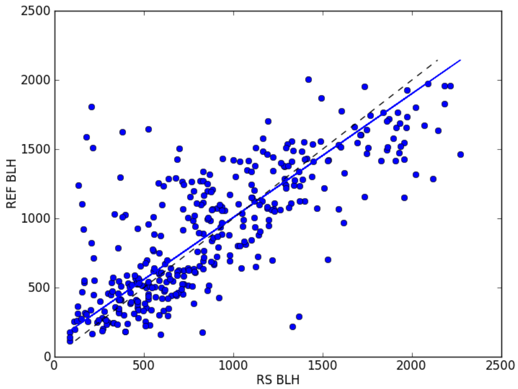
"The REF simulation captures reasonably well the BLH for all the sites with correlation coefficients between 0.89 at JUSS and 0.71 at TRN (Table 2). During daytime, biases are negative, between 8m (at TRN) and 70m (at JUSS). This can be explained for the rural site by the small negative bias on T2M (Fig.3). 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			

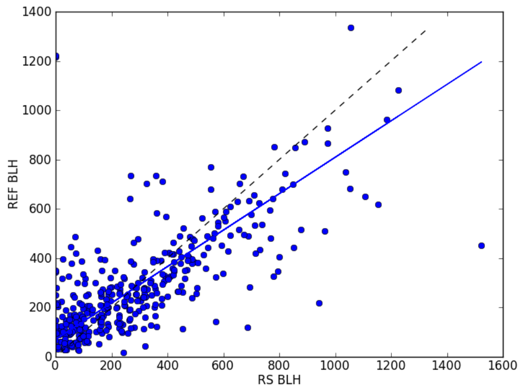
Table 2 : Statistical scores of the BLH for the REF and RUR simulations, compared to the observation. Night-time is considered from 19UTC to 8UTC.



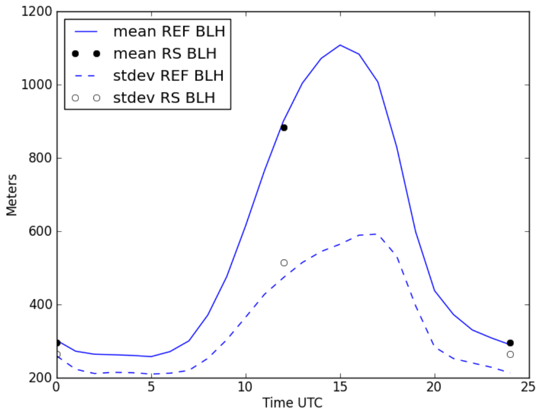
<u>Figure 6</u> : Time series of BLH (in meters above ground level (AGL)) for March 21 - 26 at JUSS (a, urban site), SIRTA (b, sub-urban site) and TRN (c, rural site).



a.



b.



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 (Tab.2), 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."

<u>Referee :</u> *From the graphs looks like that REF simulations captures better the BLH at SIRTA. A word in regard this point will be worthy.*
Line 8, page 28169. Will be better to use the term "photosynthesis" instead of "assimilation"?
<u>Authors :</u> OK

<u>Referee :</u> *Line 15, page 28169. "at EIF" repeated twice in the same sentence.*
<u>Referee :</u> *Line 25, page 28169. Adding the value of R2 would help seeing the agreement between the model and observations.*
<u>Authors :</u> Following your suggestions as well as those from the other anonymous Reviewer, quantitative assessments have been added and all the part 5 has been rewritten, as proposed 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 CO_2 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 CO_2 (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 CO_2 concentration at EIF, as the measurement is located above. Overall the statistics are worse, with a correlation coefficient of 0.05.

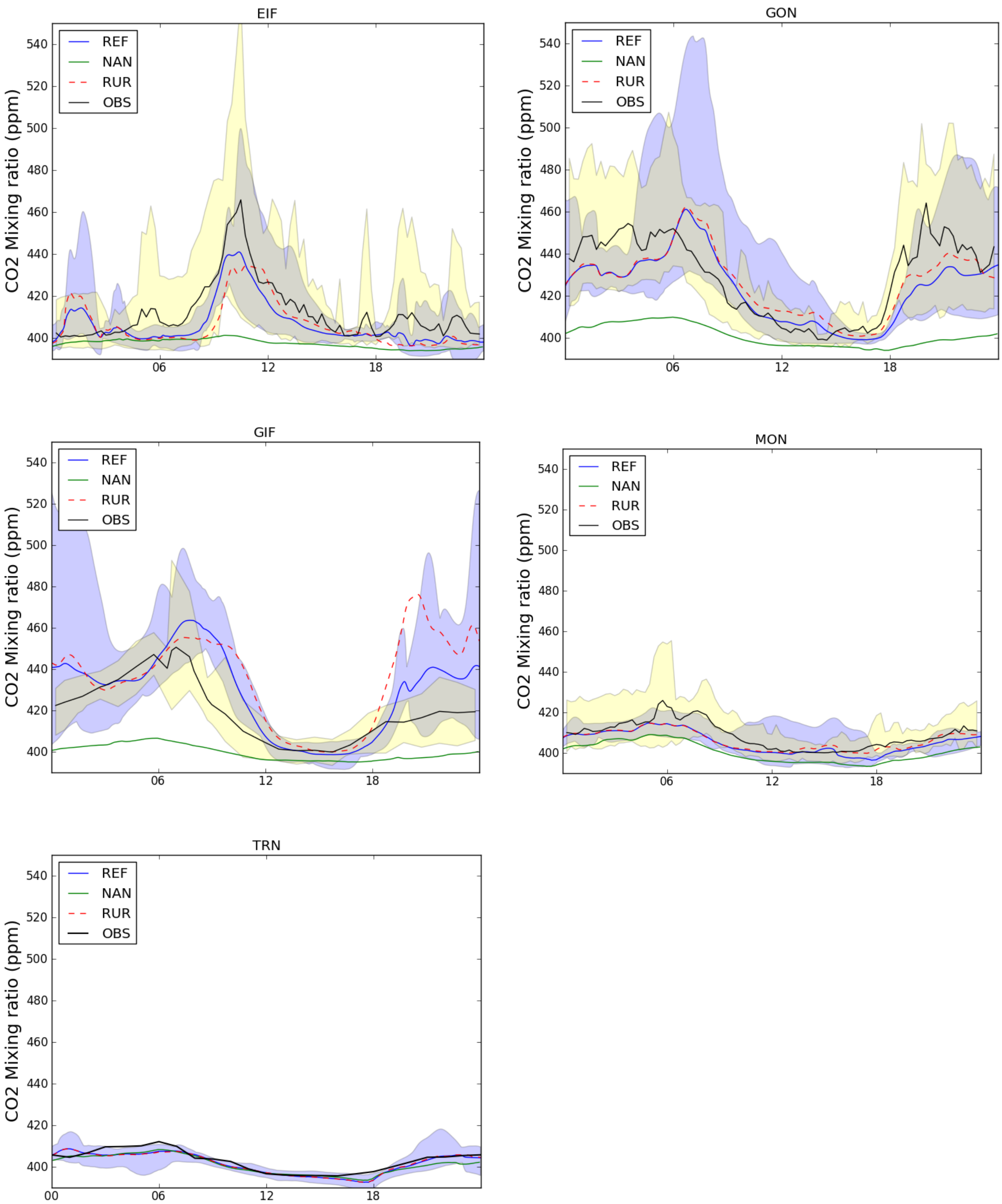


Figure 8 : Mean temporal evolution (over 6 days) of the CO2 mixing ratios (in ppm) measured (black line) and predicted with REF (blue line), RUR (red line) and NAN (green line) simulations at EIF (a), GON (b), GIF (c), MON (d) and TRN (e). The yellow area is between the minimum and maximum of the measurements,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

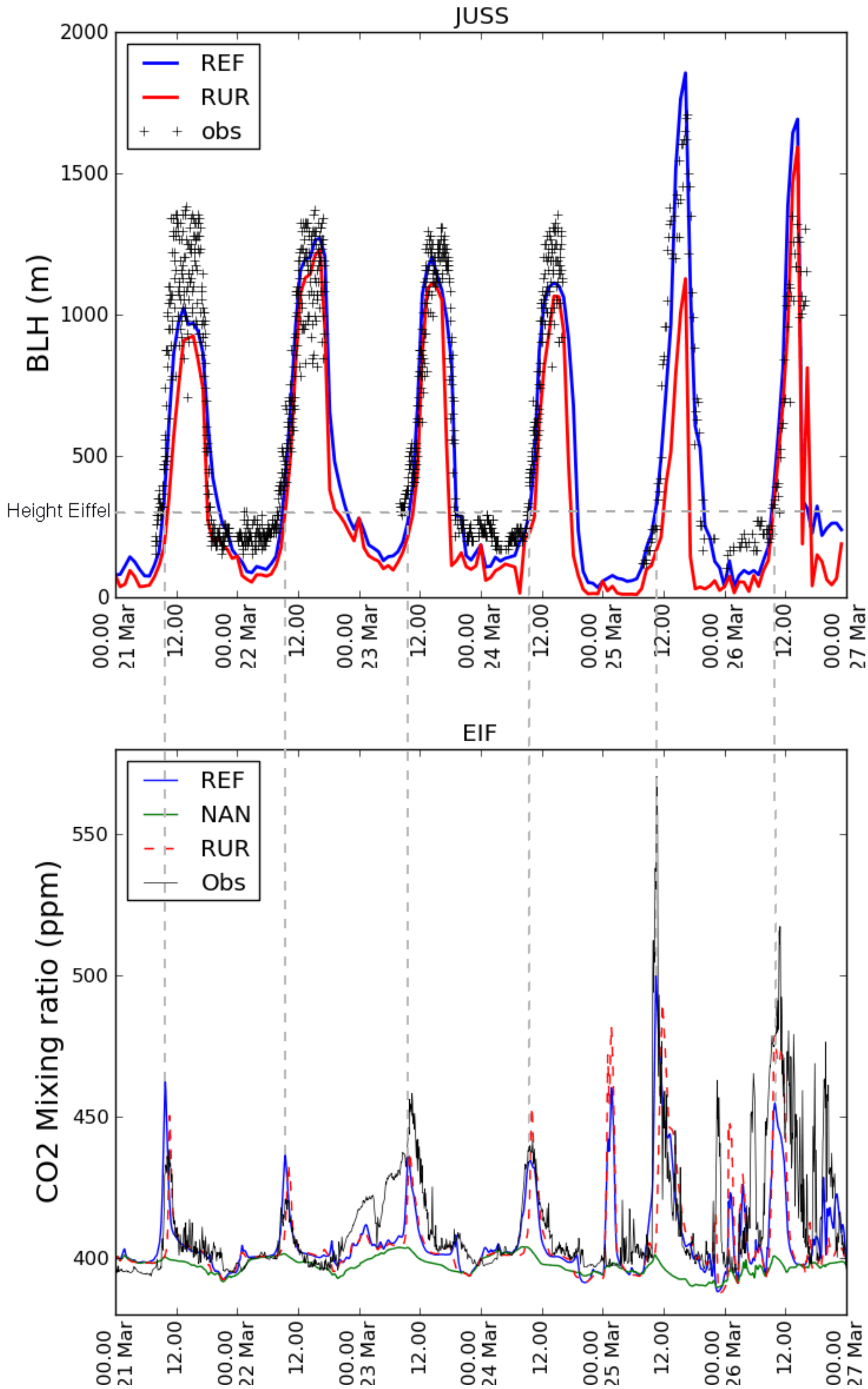


Figure 9 : (a) Time series of BLH predictions at Eiffel and observations at JUSS (in meters above ground level, AGL) for REF (blue) and RUR (red) simulations. (b) Time series of CO2 predictions and observations (in ppm) at EIF for REF (blue), RUR (red) and NAN (green) simulations. The vertical dashed lines correspond to the time in the morning at which observed BLHs reach 310m.

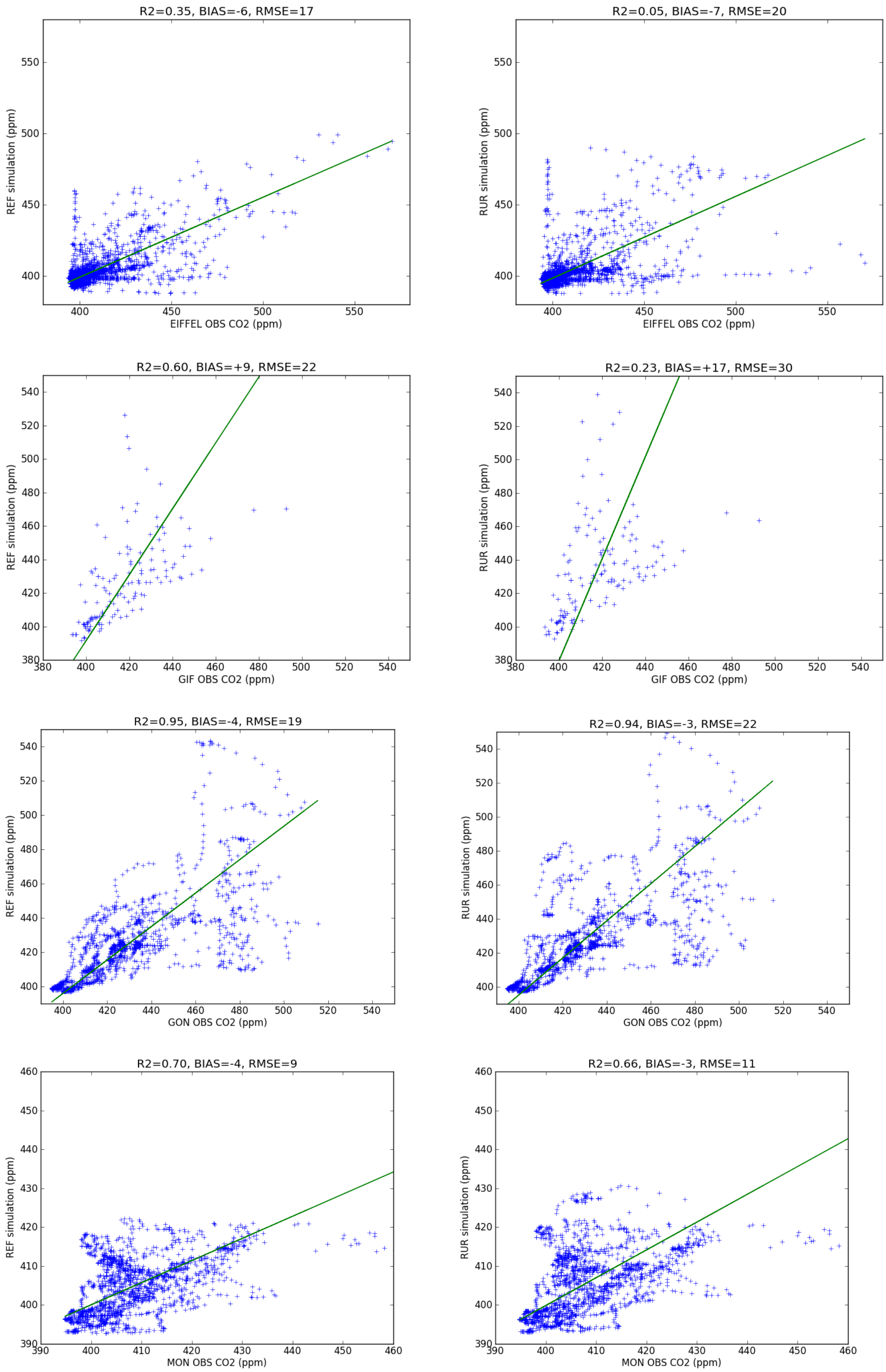


Figure 10 : Scatter plots of the observed (on the horizontal) vs. predicted CO2 mixing ratios for EIF, GIF, GON and MON for REF simulation (on the left) and RUR simulation (on the right), with an orthogonal linear regression

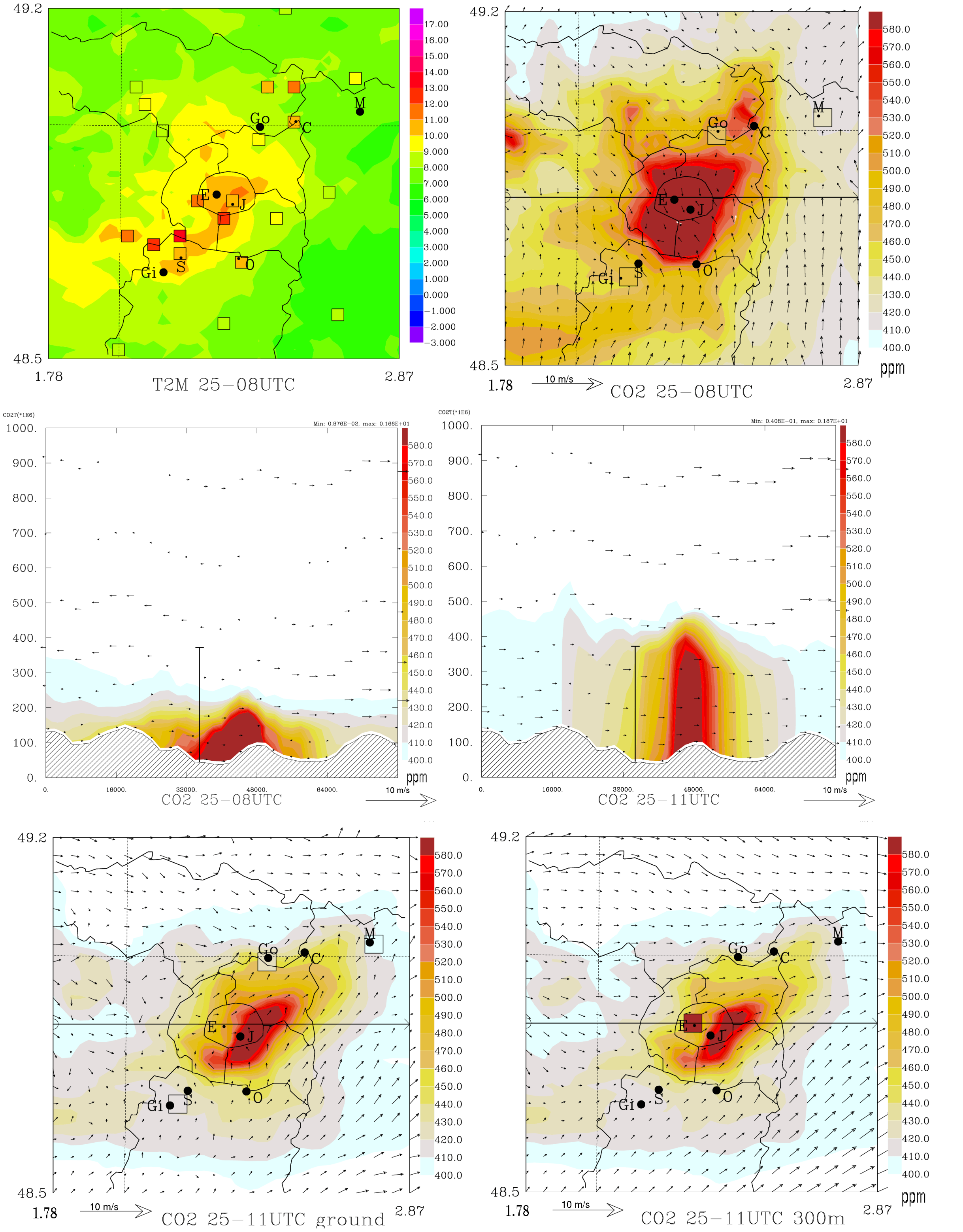
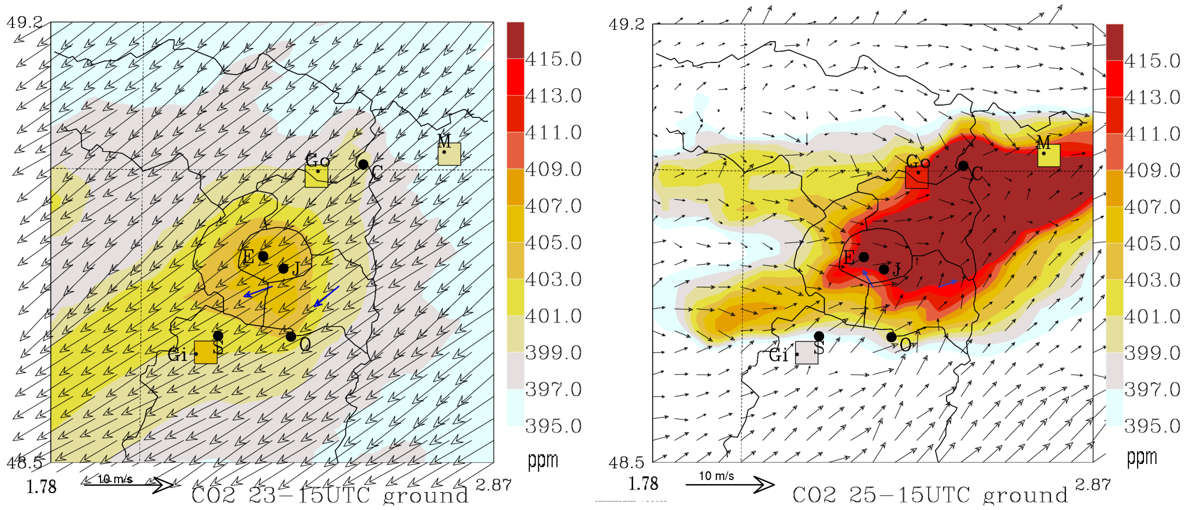


Figure 11 : MESO-NH predictions : For March 25 at 08UTC (a) 2 m temperature (in °C), (b) Horizontal cross section of CO2 mixing ratio (in ppm) near the ground. (c) Vertical cross section of CO2 mixing ratio (in ppm) according to the axis given in (b) for March 25 at 08UTC, and for March 25 at 11UTC (d), 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.

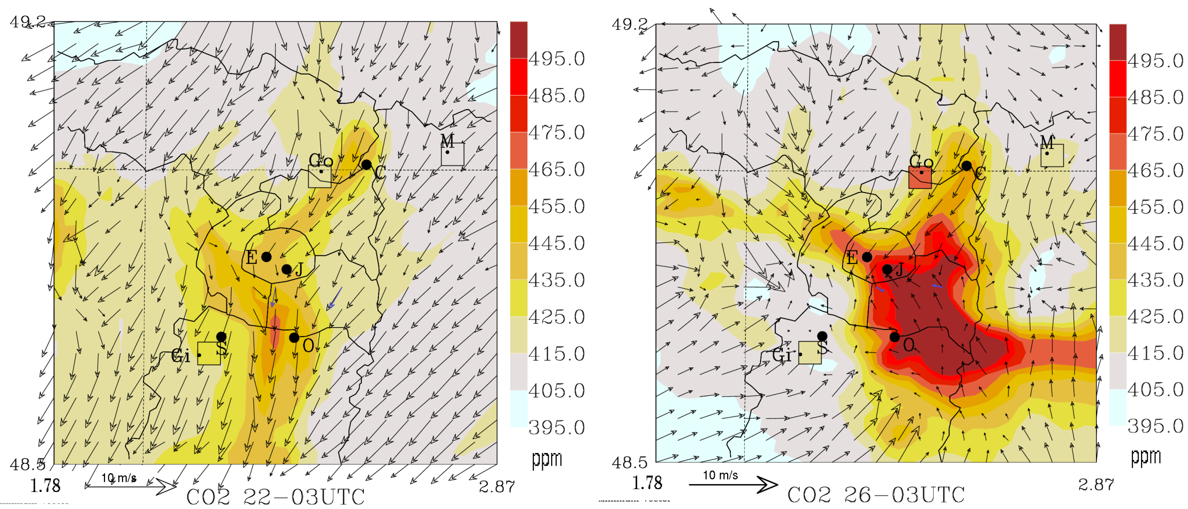
5.2 Evaluation at the sub-urban and rural sites

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.

Referee : *Line 15, page 28170. "MAGNITIUDE of the measurement". I'd say that Paris is a city rather than a town.*
Authors : OK

Referee : *Section 5.2. What does it mean "nocturnal amplitude"? Do you mean that there is an increase of 100 ppm during night? Or do you refer as "daily amplitude"*
Authors : The part concerning the amplitudes was confusing and has been removed in the new version, as based on the mean diurnal cycle.

Referee : *Section 5.2. It is confusing talking about "nocturnal peaks" and "rush hour". In most cities, rush hours take place between 7-10 am in the morning and 4-6 pm. From the graph it is difficult to assess the timing of the CO2 peaks. Does the temporal resolution of emissions inventory reproduce the rush hour peaks? If that is the case, why not trying to reduce the emissions intensity in the model and see if it reproduce better the morning rush hour peaks?*
Authors : The emissions give rush hours between 5 and 8UTC (local wintertime), and 18 and 22UTC. The temporal resolution of emissions inventory is 1 hour. Testing some modifications on the emissions is quite hazardous assuming their uncertainties, both spatially and temporally.

Referee : *Section 5.2. About the impact of CDG on GON. It is not the airport closed at night? Most of airports close between 00 and 05 local time.*

Authors : CDG 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 over 24 hours) whereas Orly has a curfew, as well as Heathrow or Francfort. CDG on the night traffic has grown faster than day. In 2011, flights heart night (0h-5h) were limited to 20 000 per year and deleted flights were postponed at the beginning and end of the night. 80% of flights at night are related to freight.

P.28169 line 9 :
"It is worth noting that peak values of anthropogenic emissions over Paris and its airports occur during rush hours, between 8 and 10UTC, and 17 and 19UTC (not shown)."

Is replaced by :
"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 near 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 over 24 hours)."

Referee : *Therefore the impact of the airport on GON measurements should be really small at night. If that it isn't the case, further evidence of the impact of the airport emissions to the data should be provided. For example, polar roses showing an increase of the CO2 concentration when the wind is blowing from that direction would be useful.*
Authors : Indeed, the impact of the airport on GON measurements is not so small at night with easterly to north-easterly winds. It is visible on the modelling fields (Fig.13 a and b) where the plume from CDG touches GON area. It is difficult to build polar roses as the duration of study is short and does not allow a climatological analysis.

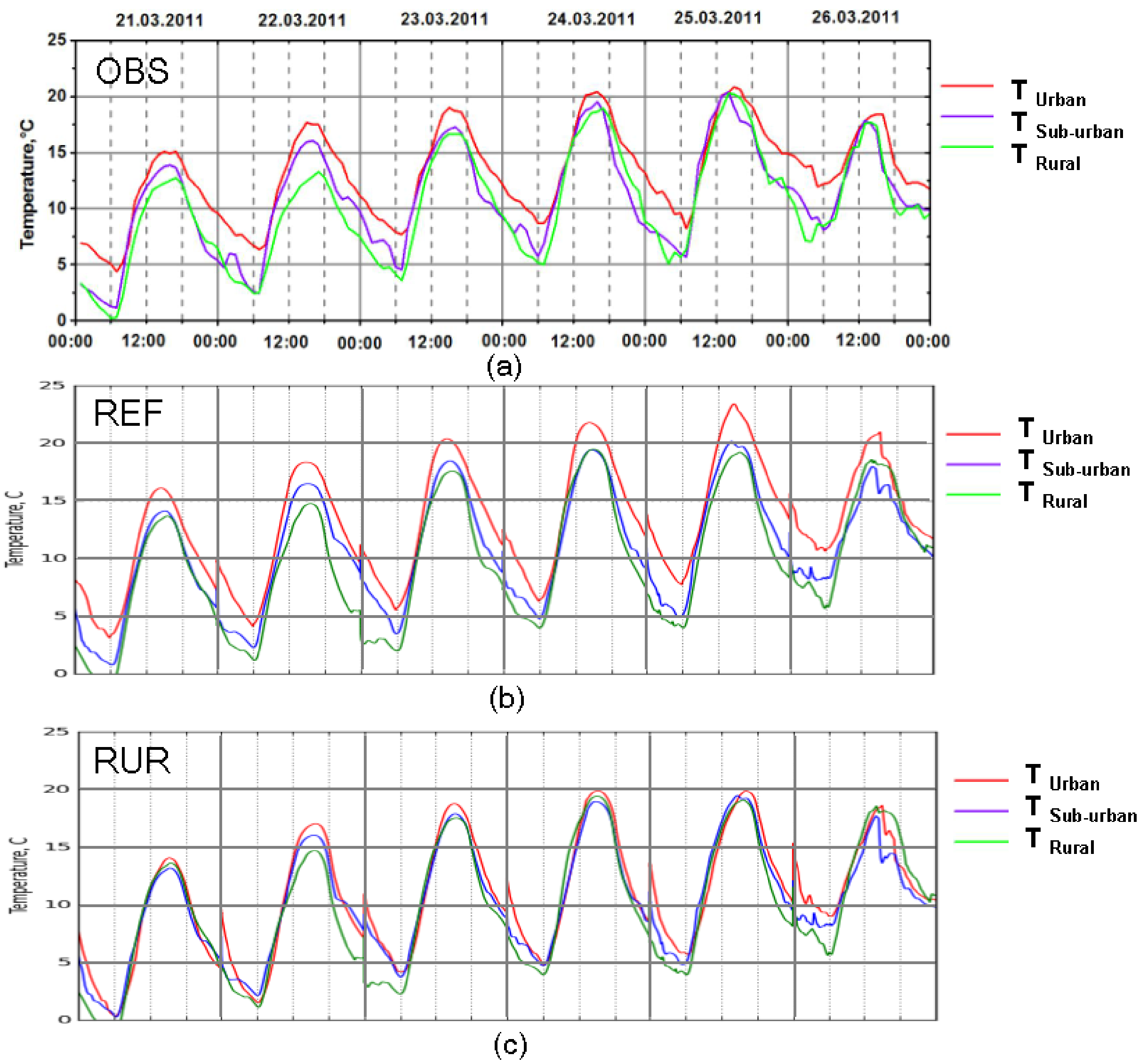
Referee : *Line 22, page 281871. I would say "EMISSIONS at rush hours are probably responsible".*
Authors : OK

Referee : *Line 23, page 28174. I don't agree that they are small discrepancies. They are reported to be between 30 and 100 ppmv, that is quite a lot. Similarly, line 3-5 page 28175 states that "the good representation of CO2 concentration on urban and sub-urban sites during nighttime" when the better agreements are found during daytime.*
Authors : OK

Referee : *Line 27, page 28174. It is pointed that the resolution of anthropogenic inventories is too coarse. Does it mean spatially or temporally?*
Authors : OK

Referee : *Fig. 3 and 7. Y-axis text for "Wind direction (degrees)"*
Authors : OK

Referee : *Fig.4 Substitute "semi-urban" in the legend by "suburban". The line colour for "suburban" is not consistent in Figs.4 b and c compared to a. In the Figure caption "predicted BY"*
Authors : This has been corrected, as well as the quality of the figure (see below).



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 CO_2 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 CO_2 diurnal cycle.

The limited duration comparison between CO_2 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 CO_2-MEGAPARIS measurement network. The study also demonstrates the potential of the CO_2-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 CO_2. However, additional CO_2 stations, especially ground stations in denser parts of the city, would be beneficial."

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 CO_2 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 CO_2 maxima occur every morning exactly at the time at which the atmospheric boundary layer (ABL) growth reaches the measurement height. The timing of the CO_2 cycle is well captured by the model, with only small biases on CO_2 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, CO_2 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 CO_2 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 CO_2 concentration at the sub-urban sites. The agreement of daytime and nighttime BLH and CO_2 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 CO_2 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 CO_2 concentrations at sub-urban sites. The diurnal CO_2 cycle is generally well captured for all the sites. At Eiffel tower, the observed spikes of CO_2 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,

CO_2 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 CO_2 concentration at the sub-urban sites. The agreement between observation and prediction for BLH and CO_2 concentrations and gradients, both day and night, demonstrates the potential of using the urban meso-scale system in the context of inverse modelling."