

Dear editor,

We apologize for the confusion. The word “reply” in bold is placed after the comment of the reviewer as it is written in the provided file. To clarify the reviewer starts with a brief repetition of the original comment then a few lines of our previous response and then continues with the updated comment. As a matter of fact we only left everything to facilitate the reading. We believe that in the new formatting it is less confusing. We also provide the page and lines (please see the new uploaded pdf file) in which we have incorporated our replies. We also kindly ask to incorporate 2 additional references (Huszar et al., 2011; Katragkou et al., 2011) in the introductory part (page 3, line 3).

Reviewer 1

2) **Original comment:** I am a bit concerned about the use of a 10km resolution meteorological model to drive a 4km resolution air-quality model...

Original reply: Some concerns: higher resolution is not the case in the study of Flagg, D.D., and Taylor suggested by the reviewer where the modelled city represents a complex multi-lake terrain.

New comment: Flagg and Taylor investigated the impact of the resolution of surface layer input data on model results for high resolution simulations of an urban area, and found that the model results were quite dependent on that resolution (e.g. root mean square differences of the heat flux between different input resolutions on the order of 20 to 30%). While both of the papers mentioned deal with cities by a coastal environment, the key issue I wanted to point out is that these models show one needs to go to very high resolution in order to capture the urban heat island circulation (Leroyer et al reference) and the effects of the surface layer changes (Flagg and Taylor reference) in an urban meteorological simulation. At the same time, I think the authors' counter-point is valid, that while higher resolution provides a better forecast in a theoretical sense, it may not do so in a practical sense. Flagg and Taylor's work implies that some of the lack of improvement may be the use of surface input data which is at lower resolution than the meteorological model, for example. As well, the downside of resolving plumes at a higher resolution is that small errors in the wind direction can result in decreases in correlation coefficient scores at higher resolution (since the lower resolution model plumes are spread out over a larger region, an error in wind direction will have less of an impact on the comparison to observations). I think the authors main arguments, once explained in the text, that they tried higher resolution meteorology in earlier work and found no improvement in the results, and that higher resolution had some additional computational overhead (though surely not as much as the air-quality model) are valid. To me, the first point suggests that more work is needed on the high resolution meteorological model - but that is beyond the scope of the current work. Discussing the issue is sufficient, here.

Reply: This discussion is added in the revised manuscript in section 2 (page 5, lines 15-31): More specifically: *“We drive the CTM over IdF using the 10km meteorology acknowledging both the computation overhead of a refined meteorology and the results of previous work over the same region using the CHIMERE CTM (Menut et al., 2005a; Valari and Menut., 2008). More specifically Menut et al. (2005a) showed that apart from the coastal areas where a refined meteorology (3km resolution) improved air-quality modelling results, in the rest of France ozone peaks were better captured with lower resolution meteorological input. Valari and Menut (2008) showed that a refined meteorological (5km resolution) input gives similar results for ozone and that model performance is much more sensitive to the resolution of emissions than to meteorology. These results suggest that in areas having the geographical characteristics of the greater Paris area (flat topography at great distance from any mountains or the ocean) increasing the resolution of the meteorological input does*

not necessarily improve the results of the chemistry-transport modelling. In contrast Flagg and Taylor (2011) investigated the impact of the resolution of surface layer input data on modelling results of high resolution simulations in an urban coastal environment and found that these results were quite dependent on that resolution. There are indications that one needs to go to very high resolution in order to capture the urban heat island circulation (LeRoyer et al. 2014) and the effects of the surface layer changes (Flagg and Taylor, 2008) in an urban meteorological simulation but this analysis is beyond the scope of the present work."

3) **Original comment:** Given the relatively small size of the meteorological and air-quality model domains, more description is needed for the downscaling and the potential impact of boundary conditions

Original reply: ...WRF simulations were carried out on a 10 km resolution grid of 90x85 cells, (i.e. 900km x 850km) which is not a "very small" domain in our opinion compared to the size of the Ile-de-France region (156km x 128km).

New comment: On the meteorology side: yes, 900x850 km is bigger than 156 x 128 km, but what about the boundary conditions used for the meteorological model (global or regional analysis)? My question there is "to what extent are the local model results affected by the driving model boundary conditions?"

Reply: Is the reviewer's question here refers to chemical boundary conditions? The sentence "*On the meteorology side: yes, 900x850 km is bigger than 156 x 128 km, but what about the boundary conditions used for the meteorological model (global or regional analysis)?*" is confusing. Does the reviewer means: "*On the meteorology side: yes, 900x850 km is bigger than 156 x 128 km, but what about the boundary conditions used for the air quality model (global or regional analysis)?*". This is answered by the reviewers suggestion in the next comment.

New comment: For the air-quality model, I was thinking of studies such as the HTAP experiments (papers by Fiore, Dentener) where they show that wintertime O3 predictions within one continent are significantly impacted by emissions changes within another continent; the latter changing the downwind continent's O3 through advection of ozone. The sentence of the authors "Having performed the simulations..." would be better stated, "While boundary conditions may impact local scale model predictions, we focus here on the impact of local emissions through the use of a common set of boundary conditions, to ensure that the differences arise from the sensitivity to local emissions."

Reply: This sentence was added in section 4.2.1 (page 16, lines 4-7).

Original comment: For the 50km resolution simulations, were the same emission data used as for the 4km resolution simulations?

Original reply: No, emission data between the regional and local scale simulations are not the same.

New comment: Again, see my slight rewording suggestion above.

Reply: To lift any possible confusion we mark that the boundary conditions to the local simulation are provided by the 0.5 degree run thus in the sensitivity analysis the only change in the 2 described simulations regards the emission inside the IdF region. Boundaries are the same. In any case the reviewer's suggestion for rewording the sentence mentioned in the previous comment is already incorporated in the manuscript.

Original comment: What boundary conditions were used for the outer 50km simulation, and where did they originate (if these were in the global coupled runs, was the model speciation the same or were there issues with matching them)?

Original reply: The matching between LMDz-OR-INCA and CHIMERE species...

New comment: The chemical table surprises me - do neither of the models include biogenic hydrocarbons (isoprene, monoterpenes)?

Reply: Yes of course. Both the LMDz-INCA and CHIMERE models include on-line pre-processing and use of biogenic VOCs. In CHIMERE this is based on the MEGAN model. In the revised manuscript we have already added in Section 2 (page 6, lines 1-12) a short description of CHIMERE which also includes a reference to MEGAN. In the LMDz-INCA model biogenics are calculated with the dynamical global vegetation model ORCHIDEE (Organizing Carbon and Hydrology in Dynamic Ecosystems). This is easily traced in the model's references.

New comment: Aside from that, what I think was needed at that point in the manuscript was a one sentence reminder to the reader that the 50km simulation boundary conditions come from a larger scale model simulation, with a slightly different chemical speciation.

Reply: This was added in Section 2 (page 6, lines 25-28).

5) **Original comment:** Some aspects of the REF versus MIT scenarios and the relationship...

Original reply: We believe that there might be a confusion regarding this issue...

This worked better. I wonder if this would be clarified further with a table with three columns going from left to right Global, Regional, Local and rows describing the different runs at each scale. It does help to have that change in the text, though.

Reply: In the revised manuscript we have added Table 1.

Original comment: Also, Figure 4 suggest that the relative impact...

Original reply: The purpose of the figure is mainly to compare each future scenario...

New comment: My point here is that the dynamic range between maximum and minimum O₃ in the rural area just outside of the IdF and within the IdF changed between the simulations; which I think is potentially interesting to mention - hence my suggestion. A few words of explanation of why the difference between rural and urban values has changed between the simulations (as opposed to focusing on the IdF in the core) would be useful.

Reply: The more drastic decrease of rural ozone under MIT is linked with the more drastic decrease of ozone precursors compared to REF and the fact that in the NO_x-limited rural environment ozone follows the fate of its precursors. In section 4.2.1 (page 16, lines 24-28) we add: *"Under the MIT scenario however, where both NO_x and NMVOCs are mitigated more effectively than in REF, ozone concentrations decrease in 2050 compared to present time levels. This feature stands-out in the NO_x-limited rural areas in which ozone follows the fate of its precursors showing a much more drastic decrease of ozone under MIT (Fig. 4c) compared to REF (Fig. 4b)."*

Original comment: The gradient between urban and rural O₃ has greatly increased in the MIT scenario and this is worth pointing out...

Original reply: If my understanding is correct this is not true actually. O₃ in the rural areas decreases much more in MIT than in REF....

New comment: Yes, but why is this the case? What in the scenarios has caused this change? My point here is that the range of maximum to minimum O₃ across the grid has changed - it's that difference that I find interesting - why have the differences between rural and urban O₃ changed between the simulations. Asking it a slightly different way: why has the rural O₃ in MIT decreased relative to the urban region much more than REF? Why has the difference between urban and rural O₃ changed between the simulations? What I was hoping for here is a sentence or two of explanation linking back to the emission scenarios employed and/or the boundary conditions.

Reply: Please refer to our reply of the previous comment.

Original comment: Page 101, line 5, a comment: Actually...

Original reply: We thank the reviewer for this suggestion...

New comment: Agreed. This raises an interesting question as to whether the variability becomes larger at local scales, hence requiring a longer averaging period. In Kelly et al, we found 7+ years seemed to get convergence. Perhaps the higher variability associated with higher resolution models requires a longer time averaging period? Something for future work, perhaps.

Reply: Indeed higher variability is associated with the increased resolution because the model can distinguish areas having very different characteristics e.g. ozone between urban, rural areas whereas a regional model of tenths of km of resolution cannot. There is a large body of literature investigating this issue (reference of Valari et al., 2008). We are currently working on a 30 year long run utilizing the same domain, emissions and meteorological forcing. We believe that the reviewer does not require some addition to the manuscript.

Original comment: Page 105: is the high bias of wind speed improved when WRF is run at higher resolution for urban regions? Given the LeRoyer et al and Flagg papers referenced above, they probably would be. See earlier comment on the resolution of the meteorological model simulations carried out.

Original reply: Yes the 10km meteorology is able to resolve better urban scale wind speeds during winter but it does have the same performance during summer.

New comment: I was referring to resolutions higher than 10km here – my question is: is there any evidence from other work that the wind speed bias improves when going to higher resolution? Perhaps in the reference quoted by the authors above?

Reply: We do not have such evidence. In Valari and Menut (2008) there is a refined meteorological simulation at 5km but the authors evaluate the differences in final concentrations and not meteorology. They show that the impact is practically non-existent. In Menut et al. (2005) a high meteorological run is employed at 3km and the authors evaluate the coarse (50km) and the fine resolution meteorology but there is no concluding evidence that the 3km meteorology is uniformly better as regards wind speed. It can predict better the sea breeze and the diurnal variation but the

intensity of wind is not much different. Again in non-coastal areas the chemical modelling results are not much different. Please we refer the reviewer to the reply of comment 2.

Original comment: Page 106, line 21: I don't follow the reasoning that short term meteorology would fail to result in 95th percentile peaks being simulated...

Original reply: This is true but episodes...

New comment: This to me suggests more a problem with the accuracy of the meteorological model rather than the time span of the simulations. The terminology used was a bit imprecise in that it allows the reader to conclude either that the time span of the simulation (short term simulation) or meteorological model is the issue here. I think that this would be clarified by the authors porting the description in the response to the reviewer into the text.

Reply: Indeed the term "short-term" is confusing and it was removed from the manuscript. A short description on the basis of our formulated answer is added in section 3.2 (page 11, line 27 to page 12, line 3): *"The 95th percentile (not shown) of observed and modeled ozone daily maxima differ by 13.8 ppb (-20.1%) indicating that the model fails to reproduce ozone levels under extreme photochemical episodes which in any case are produced in timely short periods of very specific meteorological conditions characterized by stagnated air masses and low vertical mixing favoring ozone build-up. The meteorological input used in our local simulation represents poorly the observed wind speeds which are overestimated significantly (Sect. 3.1). This affects stagnation (Jacob and Winner, 2009; Vautard et al., 2007) but also vertical diffusivity through an increased boundary layer height."*

Reviewer 2

Regarding the second reviewer we believe that our replies already reflect on the revised manuscript where it is relevant. More specifically:

Comment 1: The typical nesting used for WRF (and for CTM modelling) is a factor of 3 with consistent grid scales for both the meteorology and chemical modelling, but in this study, the meteorology is nested from 50 km to 10 km, while the CTM results are nested from 50 km to 4 km. The authors should address first how and why the meteorology is nested to 10 km while the chemistry is nested to 4 km and what are the effects of these differences.

Reply: This discussion relevant to the quality of the meteorology which is incorporated in page 5, lines 15-31.

Comment 2: Second, are there any specific issues associated with larger than normal nesting steps?

Reply: We add the next discussion in page 6, lines 17-23: *"While the nesting step is considered as larger than normal we stress that the main focus of the paper is to investigate the response of the model to the emission reductions at the local scale at a mid-21st century horizon, reductions that appear to be very strong. We show with a sensitivity on chemical regimes that local emissions is a key factor for the observed modelled changes. The nesting step from a coarse resolution of 30km or 50km a few km in the local scale is commonly used with CHIMERE (Vautard et al., 2001; Beekman and Derognat, 2003; Vautard et al., 2007; Deguillaume et al., 2008; Valari and Menut, 2008; Valari et al., 2011)."*

Comment 3: An important aspect of climate impacts on air quality is the role of biogenic emissions sensitive to temperature and other meteorological factors. There is no description of how biogenic emissions were treated and whether the biogenic emissions played any significant role in the modelled changes.

Reply: Our original reply is: CHIMERE uses an on-line coupling of the MEGAN version 2.04 model to calculate biogenic emissions. Unfortunately we cannot quantify to what extent biogenic emissions influence the observed differences. A brief discussion on qualitative basis is located on page 13, line 25 concerning the REF scenario: *“Consequently, enhanced ozone formation is expected especially in the rural part of the domain due to increase of biogenic organic compounds (BVOCs). Monoterpenes are especially sensitive to temperature while isoprene to both temperature and sunlight. We do not observe any significant changes to short-wave radiation, RH or precipitation under the REF scenario”* and on page 14, line 4 for the MIT scenario: *“Lower temperatures are expected to inhibit ozone formation while the drop of shortwave downward radiation by 16.6% relative to present will lead to less BVOCs in the rural areas.”*

Comment 4: Because a major focus in this paper is on how emissions changes affect ozone chemistry, there should be a better documentation of the changes in emission among the control and two future projection cases. A table or graphic summarizing regional and/or urban emissions (NO_x, VOCs, etc) for each case would be helpful.

Reply: We reply on why it is difficult to incorporate such table: Emission reductions for present time and the two scenarios are provided in Table 1 for NO_x, VOCs and fine particles for the local scale inventory. However, we avoided deliberately presenting any direct comparison between the local scale inventory and the GEA regional scale inventory since emission totals do not match between the two cases and such a comparison would be misleading. We believe that it is sufficient to show that air-quality projections modelled at different scales and therefore under different chemical conditions are significantly different.

Comment 5: The authors note that peak ozone episodes are not fully captured by the modelling system as evident from the current decade evaluation results. How does this aspect of model performance affect their overall conclusions related to future ozone behavior? In this case, do the results represent a lower bound on future ozone levels?

Reply: The question is answered in our original reply: We believe that failing to capture episodic ozone peaks has a negligible impact on our overall conclusion because the number of episodes in the simulated decade is too small. The underestimation of the daily maximum ozone is only ~7% in our simulation which is rather low to let us assume that future trends represent a lower bound on future ozone, especially considering that average present-time daytime values almost unbiased.