

Reviewer 1:

Specific issues:

1) On page 103, line 14, the authors mention that emissions were kept constant during the control run, and that only the vertical distribution of point source emissions across model layers varies in time. This statement can be interpreted in two ways. The first of these is that the emissions are time-invariant at all-time scales; the emission rates at every hour of every day are the same. The other is that the annual total emissions do not change between the different years of the climate run; no attempt is made to try to take into account changes in emissions levels that might take place between one year and the next.

This concern of the reviewer is simply due to poor wording. The emission inventory used in our study has full temporal variability: monthly and diurnal source-specific profiles are implemented and emissions also vary between weekdays and weekends. To avoid further misinterpretation we have now changed the corresponding text:

“The spatial allocation of emissions is completed with proxies such as high-resolution population maps, road network and the location of industrial units. It includes emissions of CO, NO_x, Non-methane Volatile Organic Compounds (NMVOCs), SO₂, PM₁₀ and PM_{2.5} with a monthly, weekly and diurnal (source specific) temporal resolution. The distribution of emissions among different activity sectors reveals that in the IdF region the principal emitter of NO_x on annual basis, is the road transport sector (50%), for NMVOCs the use of solvents (46%) and for fine particles the residential sector (40%). Annual emissions rates within the simulated decade were kept constant.”

2) I am a bit concerned about the use of a 10km resolution meteorological model to drive a 4km resolution air-quality model, particularly in an urban context. The authors' work explores the extent to which changes in resolution of emissions affect the results chemistry, but there is also a large body of work that suggests that changes in resolution of the driving meteorological model can have a significant impact on model accuracy, particularly for complex urban environments (cf. ACP special issue on the BAQSMet study, Hayden et al, Makar et al, and other papers therein, papers by Jonathan Pleim et al elsewhere, Flagg and Taylor, 2011, LeRoyer et al, 2014). WRF has been used at 4km resolution in the past – so why wasn't it used in that mode here? I would expect that the urban heat island would be better resolved, temperatures and wind velocities would be impacted, etc. The authors need to justify this in the text, or at least discuss the possible impact of using lower resolution meteorology on their results, given that they risk missing meteorological effects that could significantly affect their results. A good recent example of the impact of resolution on model meteorological predictions for urban regions can be found in LeRoyer et al, J App Met and Clim, doi: 10.1175/JAMC-D-13-0202-2013. Another: Flagg, D.D., and Taylor, P.A., ACP 11, 2954-2972, 2011.

The authors acknowledge the fact that higher resolution meteorology could possibly impact modelled pollutant concentrations and we thank the reviewer for the interesting references. However, due to the geographical characteristics of the greater Paris area (flat topography at great distance from any mountains or the ocean) and as shown in previous studies over the same region using the CHIMERE CTM (Menut et al., 2005 and Valari and Menut (2008)) increasing the resolution of the meteorological input does not improve the results of the chemistry-transport modelling. More specifically, Menut et al. (2005) showed that apart from coastal areas where refined meteorology 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 refined meteorological input gives similar for ozone and that model performance is much more sensitive to the resolution of

emissions than to meteorology. But to what extent these changes can be translated to an improvement of modelled concentrations is highly questionable. 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. Our decision to use 10km resolution meteorology was a compromise between the theoretical benefit of using fine resolution meteorological modelling and the computational cost of having to model a large domain around the city of Paris (900km x 850km) for a whole. Never the less we incorporate the above discussion in the manuscript.

Menut, L., Coll, I., Cautenet, S., Impact of meteorological data resolution on the forecasted ozone concentrations during the ESCOMPTE IOP2a and IOP2b, Atmospheric Research 74, 139 – 159, 2005.
 Valari, M., Menut, L. Does an Increase in Air Quality Models' Resolution Bring Surface Ozone Concentrations Closer to Reality?, J. Atmos. Ocean. Technol., 25, 1955-1968, 2008.

3) 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.

As far as meteorology is concerned and as mentioned in the manuscript at section 2.1. “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). As far as the air-quality modeling is concerned our domain is restricted by the bottom-up emission inventory that has been thoroughly used in both research and operational applications (Vautard et al., 2001; Beekman and Derognat, 2003; Deguillaume et al., 2008; Valari and Menut, 2008; Roustan et al., 2011) . The impact of boundary conditions is discussed in section 4.2.1, where we show that the difference in the representation of chemical regimes is mainly due to local emissions and not boundary conditions (see also our response on comment 5). We now added the following paragraph at section 4.2.1 to further discuss the potential impact of boundary conditions on our modelling: *“Having performed the simulations on the exact same domain forced by identical boundaries we also ensure that the differences that arise from this sensitivity derive from local emissions.”*

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

No, emission data between the regional and local scale simulations are not the same. The regional scale modelling uses GEA data from IIASA while our application uses local scale, bottom-up emissions. A description of how both data sets are developed is provided in Section 2.3. This difference in emission totals is the reason why we don't attempt any direct quantitative comparison between results of the 50km and the 4km simulation (see also our response to comment 5).

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)?

The matching between LMDz-OR-INCA and CHIMERE species (the last 4 lines are for particulate matter) was completed following the following table. The reader is referred to Bessagnet et al. (2009) for the exact signification of Chimere lumped SOA.

INCA	CHIMERE
Lumped >=C4 HC	n-Butane (NC4H10)
Lumped Aromatics	o-Xylene (OXYL)
C2H4	C2H4

C2H6	C2H6
C3H6	C3H6
CH2O	CH2O
CH4	CH4
CO	CO
H2O2	H2O2
HNO3	HNO3
NO2	NO2
O3	O3
PAN	PAN
DUST (CIDUSTM)	DUST (pDUST)
Sulphate (ASSO4M)	Sulphate (pSO4)
Black Carbon (AIBCM, ASBCM)	Black Carbon: 0.1-10µm (BCAR)
Primary Organic Matter (AIPOMM, ASPOMM)	Secondary Organic Aerosols (AnA1D, AnBmP, BiA1D, BiBmP) and Organic Carbon (OCAR)

Bessagnet B., L.Menut, G.Curci, A.Hodzic, B.Guillaume, C.Liousse, S.Moukhtar, B.Pun, C.Seigneur, M.Schulz, 2009, Regional modeling of carbonaceous aerosols over Europe - Focus on Secondary Organic Aerosols Journal of Atmospheric Chemistry.

The authors mention that a coupled global model was used for the initial meteorological runs – what IPCC/RCP/other emission scenarios were used for that simulation, and how did they compare to the emissions in the model simulations used here?

In Page 8, line 22 we mention that: “Two long-term scenarios are used in the global-scale simulations of future climate conditions: RCP-8.5 and RCP-2.6”. These two scenarios are used in the whole chain of meteorological simulations starting from the global and progressing to the regional (a 50km WRF simulation) and the “local” having 10km resolution. This is added additionally in Page 6, line 13.

A one-paragraph description of the CHIMERE model is needed for readers unfamiliar with the model (as opposed to just the reference) – what are the gas and particle phase components of the model and references for them, for example – just enough so that the reader has an idea as to how this model compare to others in the same class.

A paragraph is added in the relevant section (Section 2): “CHIMERE is an off-line chemistry-transport model (CTM), which models atmospheric chemistry and transport, forced by anthropogenic emissions, biogenic emissions, a meteorological simulation and boundary conditions. It is a cartesian-mesh grid model including gas-phase, solid-phase and aqueous chemistry, biogenic emissions modeling depending on meteorology with the MEGAN model (Guenther et al., 2006), dust emissions (Menut et al., 2005) and resuspension (Vautard et al., 2005). Gas-phase chemistry is based on the MELCHIOR mechanism (Lattuari, 1997) which includes more than 300 reactions of 80 gaseous species. The aerosols model species are sulfates, nitrates, ammonium, organic and black carbon and sea-salt (Bessagnet et al., 2010) and the gas-particle partitioning of the ensemble Sulfate/Nitrate/Ammonium is treated by the ISORROPIA code (Nenes et al., 1998) implemented on-line in CHIMERE”.

4) The authors use mean daytime surface O3 and Ox and the daily maximum of 8 hour running mean O3 as indicators of O3 performance and state that the model reproduce sufficiently well

both urban titration (page 105, section 3.4) and photochemical formation. These metrics will test the extent to which the model performs well for daytime urban titration and photochemical formation, but will not test how well the model performs when the titration is the dominant process, i.e., at night. The authors should also include statistics for hourly O₃, or, better, O₃ binned according to local hour, before stating that performance of titration is accurate.

Indeed we do not present night-time ozone performance in the paper. Photochemical models, including CHIMERE often fail to represent correctly night-time ozone concentrations, as shown in a number of previous studies (Van Loon et al., 2007; Vautard et al., 2007; Szopa et al., 2008). In our simulation we observe a bias of +3.5ppb during night-time. However, daytime ozone levels are well represented and given that adverse health impact is mainly correlated to daytime ozone levels we think that night-time model discrepancies should not discourage us from publishing our results. Now in section 2.2 stress this point and why our focus is on daytime ozone.

5) Some aspects of the REF versus MIT scenarios and the relationship between the regional emissions inventories and those used for high resolution were difficult to follow. If I've understood correctly from page 105, the national-level changes were used to derive local level scalings for the IdF (which makes sense, since local spatial allocation of emissions is likely to be higher resolution). However, what was done for the driving 50 km resolution simulation? This is not necessarily a trivial question – one concern I have is the extent to which the changes in the high resolution domain are due to local emissions changes versus changes in boundary conditions from the 50km model, or both. Were the emissions downscaling procedures used for the driving 50km model as well as the 4km model? This is not mentioned in the text (or I've missed it).

We believe that there might be a confusion regarding this issue. Our decade simulations are done on the 4km grid using the bottom-up inventory for the present. Indeed we use scaling coefficients from the regional inventory to derive out 2050 local emissions. The boundaries are taken from a regional simulation at 0.5° resolution which is independent from our runs, we have not intervened in the regional emissions. A second run regards the sensitivity on chemical regimes and it is completely independent from the decade simulations. In this experiment we take the emissions over IdF of the regional inventory in a particular period of a severe ozone episode and we downscale it to the 4km grid. Then we run the 4km grid with the regional emissions, we do not re-run the original 50km resolution grid. Finally we compare these results with the results of a run with the exact same configuration (resolution, boundaries etc) but we use the bottom-up emission inventory. This procedure was followed because in order to isolate the effect of emissions to the chemical regimes we needed to have also the same resolution. Re-running the native regional grid would result in differences driven both by emissions and the resolution. We understand that these differences in the description of the modelling applications could be presented clearer in the text. We utilize the term “pseudo-regional” now and not the term REG05 which is confusingly mistaken with the native regional application results. Also the impact of the boundaries on our modelling results is addressed in this sensitivity study since both the local and the “pseudo-regional” runs are driven by the same boundary conditions.

Also, Figure 4 suggest that the relative impact of local emissions is much greater in the MIT scenario than the REF scenario, but this is masked by the colour scales chosen. Figures 4(b) and 4(c) should be repeated (4d, 4e) with the same colour scale for both figures and same upper and lower limits for that scale, and no more than 10 colour bars on the scale, to show this difference.

The purpose of the figure is mainly to compare each future scenario with present time pollutant levels and to point out that ozone over Paris under MIT decreases by -3.3ppb while under REF increases by +6ppb. These points stand out in the presented figures. The Figure shows that the

relative impact of emission changes is different between the two scenarios over both Paris and the surrounding rural area. Observing the minimum values in the axes at panels b and c it becomes clear that MIT leads to largest reductions over rural areas than REF. Therefore, adding 2 more panels to the figure as suggested by the reviewer would not provide any additional information in our opinion.

The gradient between urban and rural O3 has greatly increased in the MIT scenario and this is worth pointing out, since it shows that the local emissions become much more important for some emissions scenarios than others.

If my understanding is correct this is not true actually. O3 in the rural areas decreases much more in MIT than in REF. But in REF ozone increases in Paris while in MIT ozone decreases. With a crude visual subtraction of the numbers between future and present one can see that the difference between urban-rural O3 actually becomes more or less the same amongst the 2 scenarios (~1-2ppb). So the gradient of urban-rural ozone does not increase under MIT.

6) The discussion on page 112 needs to be linked back to the earlier discussion on emissions downscaling. Are the changes discussed on lines 1 through 6 of page 112 the result of changes to absolute emissions levels and/or background emissions changes or local emissions changes? When the authors conclude at the bottom of page 112 that the regional scale emissions fail to distinguish between rural and urban chemistry, the reader is left wondering whether this is due to the downscaling procedure for the emissions, the absolute emissions levels used in the different scenarios, or both.

This last paragraph of section (Section 4.2.1) attempted to link the experiment of the sensitivity analysis of chemical regimes with the decade simulation presented through-out the paper. As mentioned in comment 5 the procedure we use isolates the effects of local (IdF) emissions. So the differences in the regimes stems from that only. Since the previous part was probably confusing this linkage attempted is not noticed. The section is re-written to be clearer.

The axes on Figure 5 should be labelled "Relative NOx", "Relative NMVOC" to underscore that these are sensitivity runs off of a starting point. Were these scalings on a mass or mole basis.

The relative changes in the axes of Figure 5 are in molecules/cm²/s; which is the corresponding input unit in CHIMERE for emissions. We have added this information in the Figure's caption. The axes have also changed according to the reviewer's suggestion.

Smaller issues:

a) Page 92, last line: another paper to add to the list of regional AQ+climate models: Kelly et al, ACP, 12, 5367-5390, 2012. Blatant plug for our own work, but it does deal with the impact of climate versus emissions that the authors mention is beyond the scope of their own work

We thank the reviewer for this interesting reference. We have added this in the revised manuscript.

b) The authors state (page 98, line 10) that regional-scale emission inventories fail to represent the plethora of emission sources at large cities. I think this needs to be qualified through some additional explanatory sentences. This may also depend on the geopolitical jurisdiction for which the emissions data have been collected, and the accuracy of spatial as well as temporal allocation data. The emissions data used for current time regional air quality models for short-term policy predictions or air-pollution forecasting are often very detailed, and

emissions processing systems such as SMOKE allow county-level data to be spatially allocated on a very local scale. Having said that, the details of those spatial and temporal allocations may be inaccurate (cf. Makar et al, GMDD, 2014) and this may lead to inaccurate model predictions. I suspect that the authors here mean to refer to “Regional scale emission inventories for future climate change scenarios”, such as the RCP scenarios described thereafter in the paragraph: those extra five words would make this clear and avoid any confusion with current climate emissions inventories (which are much more detailed than the RCP scenarios, etc)

This sentence actually refers to the level of detail of regional inventories even at present time assessments not necessarily the future. We understand that the currently used sentence by the authors is probably misleading and we have tried to remedy that. What we mean is that regional inventories cannot describe the plethora of “activities” in terms for example of the underlying technologies used within a source sector or the fact that large scale inventories use crude statistics, for example average driving speeds in a specific country’s road network. We do not consider regional scale inventories as that detailed for the finer scale at least based on European experience. We add the term “often fail to...”. The original sentence “regional scale emission inventories fail to represent the plethora of emission sources..” becomes even more relevant when we refer to future emission projections as the reviewer suggests. This part in the text is rewritten to be more explanatory.

- c) ***Page 99, line 17, “(iii) two mid-21st century:”. A few sentences describing these scenarios and placing them in the context of RCP scenarios, etc., are needed here.***

The following sentence was added to the text in the introductory part: *The “reference” (REF) scenario is consistent with long-term climate outcomes of the RCP-8.5 adopting all current and planned air quality legislation until 2030 and assumes that no climate policies are implemented thereafter. The “mitigation” (MIT) scenario is consistent with long-term climate outcomes of the RCP-2.6 and additionally assumes stringent climate mitigation policies.*

- d) ***Page 101, line 5, a comment: Actually, a 10 year run may not be that bad. The authors could check this by calculating statistics using sub-sets of their 10 year period to see the effect on the comparisons to observations, or check the standard deviations of the statistics with different numbers of years out of the total they have***

We thank the reviewer for this suggestion based on which we extracted the standard deviation for maximum ozone and daily average PM_{2.5} using different subsets of the simulated period (please see attached Figure 1). For ozone the plot shows that even if local stability regions may be found for given subsets, standard deviation varies along depending on the simulation period. For fine particles one could argue that there is a certain trend to reach stability within a 10-year simulation period but the evidence is not strong enough according to our opinion to be conclusive. Following the above, the authors believe that the original statement in the manuscript should remain.

- e) ***Page 102, lines 14 – 20. The authors seem to expect poor performance for a climate model simulating a current climate, but this should not be the case if the variables are being compared to observations on a climatological time scale (e.g. annual or 10 year averages) for both the air-quality forecast model and the air-quality climate model (e.g. Kelly et al had about the same performance for O3 for climatological and meteorological simulations with the same off-line AQ code, though worse performance for PM2.5 in the climatological simulation).***

The employed metrics (MNB,MNGE,MFB,MFE) are being calculated from daily values not from a climatological mean. We revise the paragraph in Section 2.2 (Data and metrics for model evaluation) as follows: *“We extract these metrics from the daily concentrations values and not the decade average bearing in mind that this is not typical for climate forced simulations but for operational forecast evaluation. Here we do not use re-analyses to force the regional CTM but global and regional climate simulations. Consequently one should expect lower scores based on daily values than those yielded by an air-quality forecast simulation, especially in the presence of climate biases (Colette et al., 2013; Menut et al., 2012). Since this work is original in its concept we aim to evaluate whether the urban scale setup is sufficiently realistic by utilizing metrics that are timely averaged on a scale finer than the climatological one”.*

- f) ***Page 104 lines 8 through 14: it would be useful to the reader to have a table added which gives the IdF emissions totals for major pollutants (e.g. NOx, SO2, VOCs, CO, PM2.5) for these different scenarios, to allow the reader to see the potential impact in the urban region of the different sources of information.***

Actually Table 1 facilitates this comparison. We extract emissions in a 50km² area around Paris so we are able to isolate to a large extent the urban areas from the surrounding rural areas. The only difference from the reviewer’s suggestion is that emissions are presented for the 1st model layer but the basic information on the strength of reduction is visible for ozone precursors emissions and fine particles emissions. We also note that the 50km² area around the city does not imply that emissions are that of the regional inventory. The emissions are extracted from the 4km domain, thus the future emissions in the table are the product of the hybridization methodology described in section 2.3.

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

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

- h) ***Page 106, line 21: I don’t follow the reasoning that short term meteorology would fail to result in 95th percentile peaks being simulated. If this is a climatological comparison, if the model is doing a good job and the time period is sufficiently long to be representative of the climatology, the 95th percentile peaks should be showing up in the model results (just not at the specific times in the observation record). Could the authors please clarify their argument?***

This is true but episodes (in contrast to the lower levels of ozone as well as the daily maximum in the city which is strongly linked with the variation of local emissions) are produced in very specific timely short periods and they are linked with very particular meteorological conditions (stagnation, low vertical mixing) which favour ozone build-up. Our WRF simulation does not perform a very good job when it comes to winds which are overestimated significantly. This from one hand affects stagnation and on the other hand vertical diffusivity through an increased boundary layer height leading to reduced frequency of modelled episodes as well as to milder episodes. The number of episodes (not shown in the paper) is underestimated as well. Another reason for the ozone underestimation might also be due to enhanced mixing of emissions leading to a more efficient ozone depletion. At higher model resolutions and when

NO_x gradients are strong the instantaneous mixing of emitted species at model cells can lead to large discrepancies (e.g. see Ebel et al. 2007)

Ebel, A., Memmesheimer, M. and J. Jacobs, 2007: Chemical perturbations in the planetary boundary layer and their relevance for chemistry transport modelling. *Bound.-Layer Meteor.*, 125, 265–278.

- i) ***Page 107, line 17: what is the PM_{2.5} speciation of CHIMERE (this should appear in the model description section), and what proportion of the PM_{2.5} in the IdF is primary versus secondary in origin, both in the observations and the model? Is SOA a large part of the (observed) SOA? i.e. what evidence is there to suggest that each of the factors suggested by the authors might lead to the deficit in PM_{2.5} noted? The authors go on (line 20) to state that the summertime emissions inventory might be at fault – how well does the model simulate the PBL height, vertical diffusion, etc.? These can also have a large influence on the model results.***

The modelled SOA PM_{2.5} concentrations are very low, less than 0.5% of the total wintertime concentrations and ~1.5% in summertime. This information was added in the text, although we do not have in our disposal measurements of SOA to compare. We only provide some references to justify this assumption. In particular we add in the revised manuscript 3 references by Hodzic et al. 2005; 2009; 2010, indicating that SOA are several times underpredicted by CHIMERE. The other important reasons are the particular wet climate in wintertime. This is evaluated in the manuscript and it is a known issue to the model. Another contributing factor is a possible underestimation of summertime emissions due to re-suspension which is not included in the model. There is an extensive body of literature regarding re-suspended dust in the cities and a good review with many references is Schaap et al. provided in the text. The underprediction of primary summertime emissions is referred only as a possibility that can explain a part of the underestimation. The PBL height is overestimated by 12% in the summer leading to increased vertical diffusion and lower concentrations. The latter was also added in the text.

- j) ***Page 108, line 13, “It is beyond the scope”. The Hogrefe reference already quoted and Kelly et al (ACP 12, 5367-5390, 2012) deal with the issue of climate versus emissions changes.***

By using this sentence we indicate that in this work we cannot separate the effects of climate versus pollutant emissions on future concentrations in the Ile-de-France region. We acknowledge the fact that this has been studied in different areas by other researchers as suggested by the reviewer.

- k) ***Page 111, line 18: The authors need to define in a sentence what they mean by “in-plume” chemistry here, and how they extracted it from the model. I’m assuming this means that the downwind side of the model plume from the IdF was used to generate the “in-plume” results from the rest of the paper, but this is not clear from the text.***

We now added the following lines at Section 4.2.1: *“In-plume” corresponds to the chemical reactions taking place when high concentrations of nitrous oxides coming from the city and NMVOCs from the rural areas help to accumulate ozone on the downwind site of the city. To extract “in-plume” results we identify at each hour of the simulated episode the grid cell with the maximum concentrations inside the domain.*

- l) **Page 114, section 4.2.3. The metrics being used here (Nd130, MTDM, SOMO35) may not be familiar to some readers – a short sentence or two with a reference describing their relevance as health metrics should be provided.**

We add some explanatory comments for SOMO35 at section 4.2.3: “These metrics are typically used in health impact assessment studies and account for the non-linearity in the ozone dose-response function. For example SOMO35 was developed in the Joint WHO/Convention Task Force in 2004 and represents the cumulative annual exposure to ozone. The threshold value was set based on evidence that a statistically significant increase in mortality risk estimates was observed at ozone concentrations above 25–35ppb”.

- m) **There are several cases where there are minor errors in the use of English**

We tried to revise the manuscript thoroughly for errors in the use of English.

Reviewer 2:

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.

Our reasoning behind our decision to stay at 10km resolution was based on work done with the same model (CHIMERE) in France. First the authors acknowledge the fact that a refined meteorological resolution can impact the quality of the simulations, but given the fact that Paris is characterized by a rather flat topography, away from the ocean, higher resolution meteorology is not expected to improve the results of the air-quality modelling at 4km. More specifically, Menut et al. (2005) showed that apart from coastal areas where refined meteorology 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 input gives similar results for ozone and that model performance is much more sensitive to the resolution of emissions than to meteorology. This discussion as regards WRF is incorporated in the manuscript. From the WRF official website (http://www.mmm.ucar.edu/wrf/users/docs/user_guide/users_guide_chap5.html#_e_One-way_Nested) we quote: “when one-way nesting is used, the coarse-to-fine grid ratio is only restricted to be an integer. An integer less than or equal to 5 is recommended”.

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

The main focus of the paper is the response of the model to the emission changes at local scale. We believe that the emissions strength reductions in the future are sufficiently large to outbalance any difference that would be caused by more “reliable” boundaries provided by an intermediate domain. The nesting step from 30km or 50km to a few km of resolution 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). In any case in the sensitivity analysis of chemical regimes we show that local emissions is a key factor to the observed change of the regime and can explain the response of the area to the future emissions reductions. Apart from some influence in the rural stations that are closer to the boundaries we do not consider that an intermediate step in the CTM would produce something significantly different in the high-titrating areas.

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.

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 108, line 19 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 109, line 6 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.”

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 (NOX, VOCs, etc) for each case would be helpful.

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.

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 behaviour? In this case, do the results represent a lower bound on future ozone levels?

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.