Anonymous Referee #2
Received and published: 17 November 2015

In this work we are introduced to a methodology, which is based on the use of a simple statistical model, which aims to confirm the impact of climate change impact on air quality and provide a range of uncertainty. The study is focusing on PM2.5 and ozone. While I understand the motivation for the development of such a handy tool, which will allow a fast and low-cost assessment of the impact of climate change on air quality, I am very skeptical on the methodology followed.

My major concern is the use of a very simplistic statistical model which tries to relate linearly PM2.5 with selected meteorological parameters. Considering that PM2.5 consist of a number of chemical species, which have various dependences on different meteorological variables (depending on their physical and chemical characteristics e.g. hygroscopicity, optical properties etc), I think it is not very responsible to assume a linear link between meteorology and PM2.5. My suggestion is that authors focus on ozone data only, and omit completely the analysis of PM2.5. Therefore, hereafter my comments will focus only on findings related to O3.

We understand the concern of the reviewer about the complexity of ozone chemistry and the fact that PM2.5 consists of numerous chemical species aggregated.

An important reason why the model we propose has skill is that we focus here on aggregated quantities (spatial averages, daily means), whereas non-linearities would have been much larger at high spatial resolution and temporal frequency. There are analogous findings in the study from (Thunis et al., 2015) who demonstrated that annual mean ozone and particulate matter responses to incremental emission changes were much more linear than previously thought. The following has been added in Section 2.1 to better explain this issue (page 6 line 21):

> "It should be noted that focusing on aggregated quantities greatly improves the skill of the statistical model that would struggle in capturing higher temporal frequency and spatial resolution. An analogy is presented in (Thunis et al., 2015) who demonstrated that annual mean ozone and particulate matter responses to incremental emission changes were much more linear than previously thought."

We also took the opportunity of this review to explore more sophisticated Generalized Additive Models as explained in more details when answering to the specific related comment below.

In order to address the Reviewer concern about possible compensating effects when using statistical regression for the PM2.5 mix of constituents, these more sophisticated statistical models were applied to each sub-constituent contributing to the PM2.5 mix and the corresponding results are presented in new Section 3.2.2 of the revised manuscript. That reads:

> "Because total PM2.5 is constituted by a mix of various aerosol species, there is a risk of compensation of opposite factors in the statistical model. In order to assess that risk, we developed such models for each individual PM constituent in the chemistry-transport model. The performances of these statistical models in terms of correlation for the historical (training) period or in predictive mode for the future period (testing) are presented in Figure 3.
>
> For all regions, the statistical models are not able to capture the variability of mineral dust. This is because the design of the statistical model is exclusively local (i.e. average concentrations over a given region are related to average meteorological variables over the same region), whereas most of the mineral dust over any European region is advected from the boundaries of the domain, in North Africa. It should be noted however, that except for the regions IP and MD, the dust represents only a small fraction of the PM concentrations (Figure 4). That could explain why the statistical model for PM2.5 performs poorly over IP and MD, but it will not undermine the confidence we can have in concluding about the robustness of the PM2.5 model for the region selected above: ME, EA and NI.
>
> All over Europe, primary particulate matter (PPM) is one of the smallest particulate matter fractions. Their variability is well captured by the statistical model for all the regions except SC. But because of their small abundance in that region, they should have a limited impact on the PM2.5 model performance.
>
> The sea salts are well reproduced by the statistical model for all the regions except NI and EA. These two regions have no maritime area, therefore sea-salt concentrations are lower and exclusively due to

advection which, as a non-local factor, is not well captured by the statistical model.

Ammonium (NH_4^+) aerosols are satisfactory captured by the statistical models for five regions out of eight including those selected for the overall PM2.5 model (ME, EA and NI).

The organic aerosol fraction (ORG) is well reproduced over the historical period and the predictive skill is satisfactory (NRMSE around 0.7) for ME, EA and NI.

The statistical models are efficient to reproduce the nitrate (NO_3^-) concentrations over the historical period for ME, EA, AL, MD, FR & BI regions but the predictive skills are only considered satisfactory for ME, EA, FR and NI, where nitrate constitutes a large fraction of PM2.5.

Sulphate aerosols (SO_4^{2-}) are well represented by the statistical models for BI, EA and ME. The performances are low in the NI region, but sulphates constitute one of the smallest particulate matter fractions for that region.

This analysis of the skill of statistical models for each compound of the particulate matter mix confirms that there is no compensation of opposite factors in the selection of skillful models for total PM2.5 proposed in Section 3.2.1. The only cases were one of the particulate matter compound was not well captured by a statistical model, could be attributed to a low, and often non-local contribution of the relevant particulate matter constituent for the considered regions. We conclude that the selection of ME, EA and NI as regions where it is possible to build a statistical model of PM2.5 variability using Generalized Additive Models based on meteorological predictants would hold if the model had been built for each constituent of the particulate matter mix."

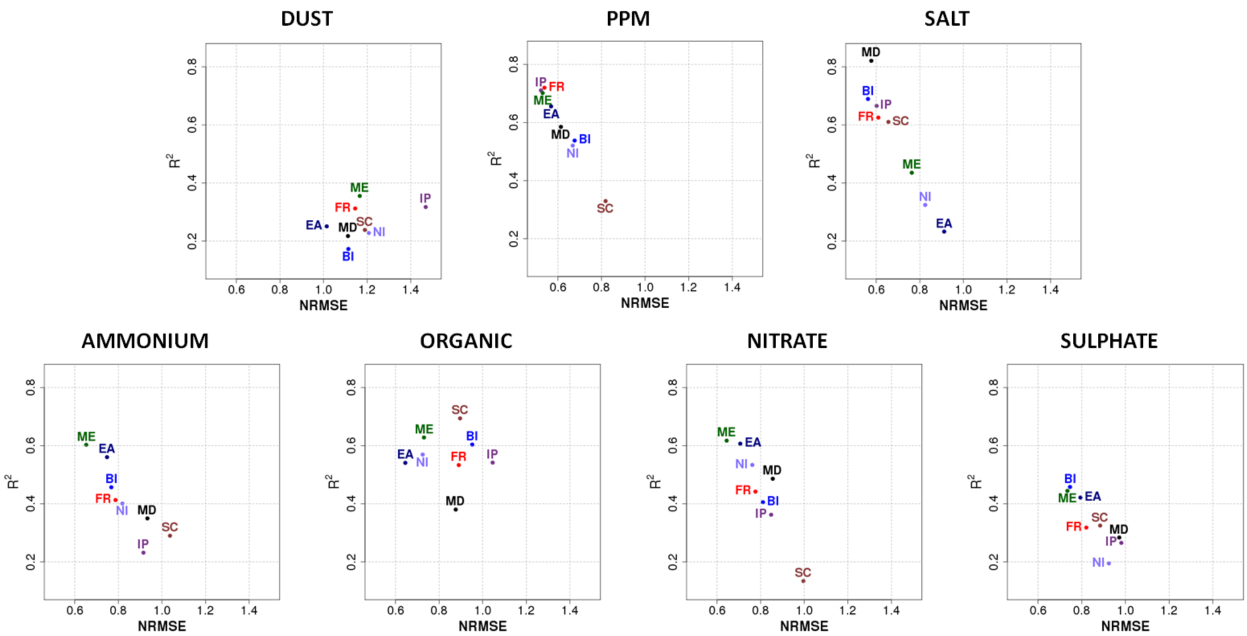

Figure 3: Statistical model evaluation for each particulate matter constituent (from left to right: PM2.5, Dust, Primary Particulate Matter, Sea-salt, Ammonium, Organic fraction, Nitrate, Sulphate). The x-axis represents the Normalized Mean Square Error applied to the delta (future minus historical) of either the generalized additive model or CHIMERE. The y-axis represents the R² of the statistical model (training period).

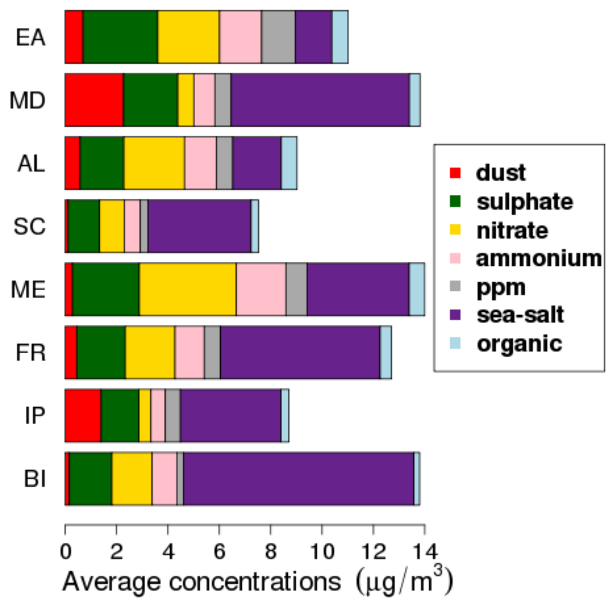

Figure 4: Average particulate matter composition for the historical period per region.

Because of the reasons explained above, we chose not to exclude PM2.5 from the analysis.

Throughout the manuscript, authors claim that with the use of their simple statistical model they can "conclude on the robustness of the climate impact on air quality". I think this statement exaggerates on what we can expect from a simple statistical model, the use of which, in my opinion, is to provide quick-looks or quick estimates, in a fast and low-cost manner. I would rather tend to consider as "conclusive", a result from a sophisticated numerical model, or even better, as authors also mention, results of several ensemble members. Therefore, I would suggest the rephrasing of relevant sentences in the revised manuscript.

We rephrased the relevant sections, pointing out that such conclusions on the robustness (i) are limited by the skill of the statistical model, therefore only valid for selected regions and species, (ii) should be further tested by implementing full CTMs with the subset of RCMs identified as priorities.

We would like to emphasize that the supposed better quality of more sophisticated numerical models can at present only be reached at the cost of using a very limited number of regional climate projections (one in most studies, at best two in a couple of studies), therefore leaving a high risk in not offering an appropriate coverage of climate uncertainties. Our approach therefore offers a complementary assessment of such uncertainties, as now better explained in the introduction (page 3, line 31):

> "This method allows selecting the members of the RCM ensemble that offer the widest range in terms of air quality response, which could be considered as air quality sensitivity to climate change projections. These selected members should be used in priority in future air quality projections. A byproduct of our statistical air quality projections is that we explore an unprecedented range of climate uncertainty compared to the published literature that relies, at best, on two distinct climate forcings. The confidence we can have in these statistical projections is of course limited by the skill of the statistical model. Our approach of using a simplified air quality impact model but with a larger range of climate forcing can therefore be considered complementary with the more complex CTMs used with a limited number of climate forcings."

I would strongly suggest the use of a more sophisticated statistical model, instead of the proposed linear model, to assess the impact of climate on O3. Especially if results prove to be different than those presented in the current analysis (i.e. if current findings are not reproducible), we can be sure that the current methodology suffers from several caveats –which we already known and authors already addressed - which cannot be ignored.

Building upon the advice of the referee, we have investigated the substitution of Linear Model by Generalized

Additive Models (GAM) and also included a focus on each component of PM in Section 3.2.2.

We should however point out that none of the general conclusion of the original manuscript in terms of the climate penalty/benefit on ozone/PM is changed, and the RCM selected as a priority for further analysis are the same, thereby demonstrating the robustness of the approach itself.

"4.1 Fine particulate matter

In order to assess qualitatively the robustness of the evolution of regional climate variables having an impact on air quality, we first design a 2-D parameter space where the isopleths of statistically predicted pollutant concentrations are displayed (background of Figure 5). Then the distributions of historical and future meteorological variables as extracted from the regional climate projections are added to this parameter space. For each Regional Climate Projection, we show the average of the two driving meteorological variables as well as the 70th percentile of their 2D-density plot, i.e. the truncation at the 70^{th} quantile of their bi-histogram which means that 70% of the simulated days lie within the contour. Both historical and future climate projections (here for the RCP8.5 scenario and the 2071-2100 period) are displayed on the parameter space. The climate projections are all centered on the IPSL-CM5A-MR/WRF member so that only the distribution of the latter is shown for the historical period.

As pointed out in Table 1, the main meteorological drivers are the depth of the PBL and near surface temperature for the example of PM2.5 over Eastern Europe region displayed in Figure 5. The statistically modeled isopleths in the background of the figure show that PM2.5 concentration decrease when the depth of the PBL increases (x-axis), or when temperatures increase (y-axis). The interactions captured by the GAM exhibit the strong influence of high vertical stability events (with low surface temperature and PBL depth) in increasing PM2.5 concentrations. On the contrary, for high temperature ranges, the depth of the PBL becomes a less discriminating factor. The comparison of historical and future distributions shows that both meteorological drivers evolve significantly in statistical terms (Student t-test with Welch variant at the 95% confidence level based on annual mean). However, even though the PBL depth constitutes the most important meteorological driver for PM2.5, it does not evolve notably compared to the surface temperature in the future (Figure 5). Thus the largest increase of the secondary driver (surface temperature) leads to a decrease of PM2.5 concentrations. The largest and the smallest PM2.5 concentrations decrease are found for CSIRO-Mk3-6-0/RCA4 and MPI-ESM-LR/CCLM, respectively. But the overall spread of RCMs in terms of both the evolution of PBL depth and temperature is limited, suggesting that this climate benefit on particulate pollution is a robust feature. Those isopleths present the same characteristics for ME and NI regions (Supplementary information Figures S1, S4). The qualitative evolution represented in Figure 5 is further quantified by applying the GAM to the future meteorological variables in the regional climate projections. These results are represented by the probability density functions of the predicted concentrations of each GCM/RCM couple minus the estimated values for the historical simulation (e.g. 2071-2100 vs. 1976-2005, Figure 6). For EA and ME, the longer tail of the probability density function of MPI-ESM-LR/CCLM compared to the average of the models reflects that stronger pollution episodes will occur in the future even if the mean of the concentrations is lower than the average of the ensemble (Figure 6 for EA and Figure S2 for ME).

Besides the distribution, the ensemble mean and standard deviation of the estimated projected change in PM2.5 concentrations has been quantified (Table 3). All the selected regions depict a significant decrease of the PM2.5 concentrations across the multi-model proxy ensemble indicating that according to the GAM model, the climate benefit on particulate matter is a robust feature in these regions. The magnitude of the decrease depends on the region, its ensemble mean (± standard deviation) is -1.08 (± 0.21) $\mu g/m^3$, -1.03 (± 0.32) $\mu g/m^3$, -0.83 ± (0.14) $\mu g/m^3$, for respectively EA, ME and NI (Table 3).

In order to explain the differences in the response of individual RCM in the ensemble, we need to explore the historical meteorological variables probability density functions (PDF, Figure 7) and to compare them with the evolution of IPSL-CM5A-MR/WRF (Figure 5). The comparison of the historical distribution for the temperature reflects the stronger extremes of IPSL-CM5A-MR/WRF (e.g. colder than the others when it is cold). It is only for the NI region that IPSL-CM5A-MR/WRF lies in the mean of the ensemble. Concerning the PBL depth, the values are similar to the average of the ensemble for ME

even if MPI-ESM-LR/RCA4 and EC-EARTH/RACMO2 present the largest values. IPSL-CM5A-MR/WRF has a thinner boundary layer for NI and a deeper than the average for EA but the differences are limited Figure 7).
It is for CSIRO-Mk3-6-0/RCA4 that we find the most important decrease of PM2.5 for the selected regions (Table 3). This is related to a larger temperature rise compared to the other models and a larger boundary layer height increase compared to the other member of the ensemble for these regions Figure 5). CanESM2/RCA4 and CSIRO-Mk3-6-0/RCA4 exhibit the same features for the ME region.
MPI-ESM-LR/CCLM presents the smallest decrease of PM2.5 for each of the selected regions (e.g. over ME is almost 3 times smaller than the largest decrease) except EA where CNRM-CM5-LR/RCA4 presents a smaller decrease (-0.77 μg/m^3 vs. -0.81 μg/m^3). As already mentioned above, the particular tails of the statistically modelled PM2.5 distributions for EA and ME indicate a larger contribution of large pollution episodes in the future for that RCM. But the historical distributions exhibit a larger boundary layer than the average models of the ensemble and a similar temperature. Thus, the low PM2.5 concentration decrease is explained by the limited average evolution of the meteorological drivers as shown in Figure 5.
Overall we conclude that a climate benefit is identified for the PM2.5 for each of the selected regions. To the extent that the statistical model is skillful, as demonstrated in Section 3.2.1, this result is robust across the range of available climate forcings since the whole ensemble of regional climate projection present consistent features. The regional climate models that exhibit the largest and smallest responses are CanESM2/RCA4; CSIRO-Mk3-6-0/RCA4 and MPI-ESM-LR/CCLM, which should therefore be considered in priority for further evaluation using explicit deterministic projections involving full-frame regional climate and chemistry models.

4.2 Ozone peaks

For most of the selected regions (FR, IP, ME and NI,) the main drivers are the same (i.e. near surface temperature and short wave radiation). The isopleth in the background of Figure 5 show that temperature and short wave radiation have a similar impact on ozone peaks, except in the larger range of short wave radiation anomalies, where temperature becomes less discriminating. All the isopleths (Figure 5 for EA and Figures S1, S4 and S7 for ME, NI, FR and IP) exhibit an increase in the distribution of temperatures because the projected future is warmer than the historical period. According to the ozone peak concentrations predicted by the GAM (displayed in the background of Figure 5) these increases will lead to more ozone episodes. This trend appears for the entire models ensemble so that we can conclude with confidence that the climate penalty bearing upon ozone is a robust feature even if the specific distribution of some of the models stand out (CanESM2/RCA4; CNRM-CM5-LR/RCA4; CSIRO-Mk3-6-0/RCA4; IPSL-CM5A-MR/WRF).
The ozone increase of the ensemble reaches +10.51 (± 3.06) μg/m^3, +11.70 (± 3.63) μg/m^3, +11.53 (± 1.55) μg/m^3, +9.86 (± 4.41) μg/m^3, +4.82 (± 1.79) μg/m^3 for EA, FR, IP, ME and NI Table 3). These values confirm the statistically significant climate penalty (the mean is at least two times larger than the standard deviation). However, as already mentioned for Figure 5, we find minor differences among the models. The meteorological distributions are marginally different between the models of the ensemble: the summertime temperature predicted by IPSL-CM5A-MR/WRF has stronger extremes than the other models. Moreover, it is warmer than the ensemble in EA. Concerning incoming short wave radiation, IPSL-CM5A-MR/WRF lies in the average (Figure S3, S6, S9) except for the region EA where the amount of incoming radiation is the highest among the ensemble (Figure 7). Note that, only EC-EARTH/RACMO2 and MPI-ESM-LR/RCA4 exhibits lower values (around half of the average for MPI-ESM-LR/CCLM). The lower amount of summertime incoming short wave radiation for the couple MPI-ESM-LR/CCLM is relevant for all the selected regions.
The magnitude of the ozone rise differs between the models and the regions. Note that CanESM2/RCA4 exhibits the largest difference (i.e. around 1.5 times the ensemble mean) followed by CSIRO-Mk3-6-0/RCA4 for each selected regions. This is explained by the larger temperature increase during summertime which is the major driver, as identified by the statistical models, of ozone concentration. Note that the value is skyrocketing for the region ME, 5 times the value of IPSL-CM5A-MR/WRF which shows one of the lowest increases. CNRM-CM5-LR/RCA4 presents the lowest increase.
On the contrary, the lower increase of the summer temperature and sometimes a decrease of the

incoming short wave radiation amount (e.g. IPSL-CM5A-MR/WRF in NI) lead to lower ozone concentration changes for IPSL-CM5A-MR/WRF and CNRM-CM5-LR/RCA4 for FR, IP, ME and NI (Table 3). Note the specific evolution for the region NI, where the IPSL-CM5A-MR/WRF model yields almost no increase of the ozone concentration compared to the other models while on the map of the differences in the deterministic model (Figure 1.f), the evolution was statistically significant. This absence of evolution reflects the limitation of the statistical models.
In figure S5, we can point out an outstanding pattern of the MPI-ESM-LR/CCLM distribution for the NI region with particularly large tails. The ozone rise would be more pronounced for the upper quantile which depicts more extreme ozone pollution episode (note that this was also the case for that model in terms of PM2.5 pollution).
Overall the climate penalty is confirmed even if some regional climate models stand out of the distribution, such as CanESM2/RCA4; CNRM-CM5-LR/RCA4 and CSIRO-Mk3-6-0/RCA4 which should therefore be considered for further deterministic projections."

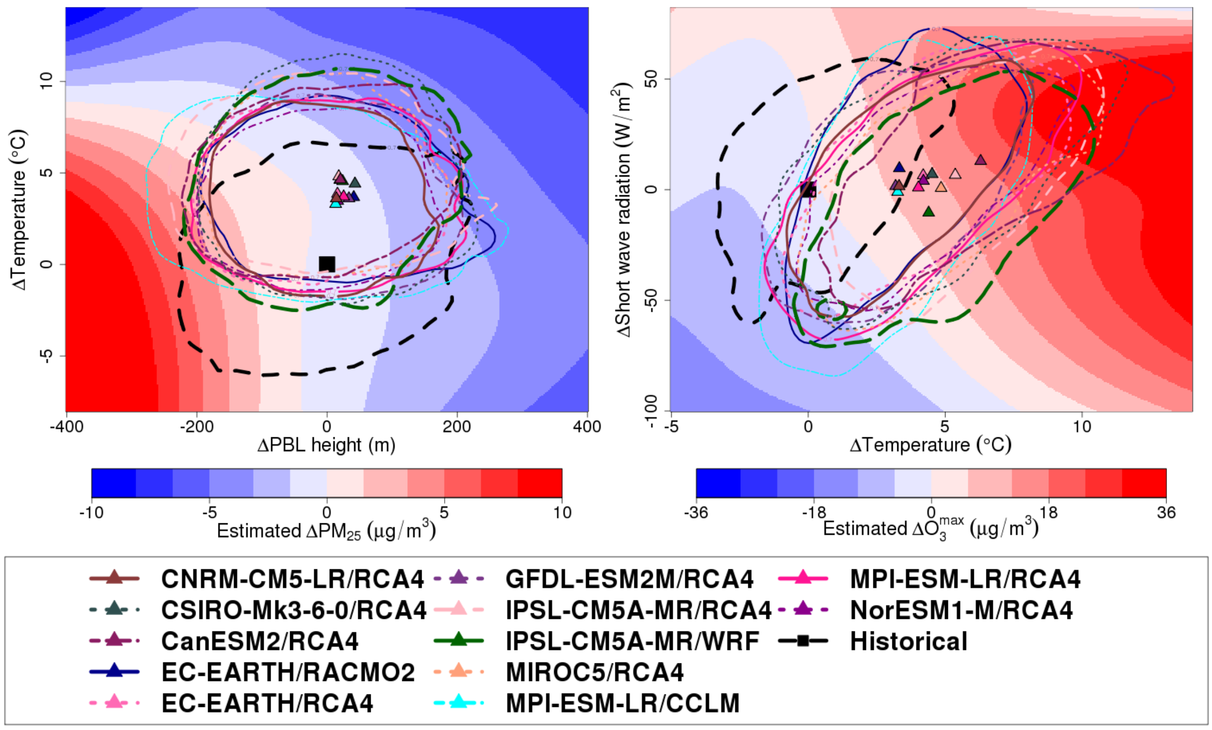

Figure 5: The left figure presents the proxy of ensemble projections for daily average de-seasonnalised PM2.5 concentrations in Eastern Europe. The right figure represents the proxy for daily maximum de-seasonnalised summer ozone for Eastern Europe. For both figures, the shaded background represents the evolution of pollutants estimated by the statistical models. The contours are representing the regional climate projections and the triangles their mean. The black dashed contour represents the historical – IPSL-CM5A-MR/WRF – and the square its mean.

We are not aware of previous studies highlighting caveats in the approach we propose, in particular because we focus on long term changes for aggregated indicators (regional averages and daily means), while it would indeed be more challenging to capture high-frequency changes at a monitor location for instance.

Another major concern comes from the resolution of the regional models used. A 50-km resolution is not recommended for impact studies. One cannot really expect much from the impact-assessment point-of-view in such a coarse resolution. Especially when higher resolution simulations are available in the community, what justifies the selection of 50 Km resolution model results? I would strongly suggest the use of 12 Km resolution simulations.
The appropriate model resolution differs with the climate impact being considered. As far as climate impacts on air quality are concerned, a higher resolution in the CTM is probably desirable to better address population

exposure, even if it has been demonstrated that this improvement lies more in the refinement of emission inventories rather than meteorological (climate in our context) forcing (Valari and Menut, 2008).

Using a higher resolution climate forcing has been shown to improve the representation of extremes such as daily precipitation intensity (Jacob et al., 2014). However we should point out that air quality is not sensitive to precipitation extremes (triggering of low precipitation events, or blocking situations would be much more sensitive). In addition, again we limit the analysis to aggregated European regions, and (Kotlarski et al., 2014) demonstrated that for aggregated quantities over such subdomain, no apparent benefits of a finer grid are identified.

For air quality projection in a future climate context, 50km is the state of the art, none of the numerous papers published, even very recently, is using a higher resolution (Meleux et al., 2007;Langner et al., 2012a;Langner et al., 2012b;Manders et al., 2012;Colette et al., 2013;Hedegaard et al., 2013;Watson et al., 2015).

According to the reviewer comment, we add the following paragraph (page 7, line 15) to discuss the choice of such a resolution:

> "Whereas higher spatial resolution simulations are available in the EuroCordex ensemble, the 0.44 resolution were considered appropriate for air quality projections in agreement with other publications (Meleux et al., 2007;Langner et al., 2012a;Langner et al., 2012b;Manders et al., 2012;Colette et al., 2013;Hedegaard et al., 2013;Watson et al., 2015), and also because higher RCM resolution are not specifically performed to improve the climate features that are most sensitive for air quality purposes (temperature, solar radiation, stagnation events, triggering of low-intensity precipitation events etc.)."

However, if this is not possible, authors should definitely go into a very detailed description of evaluation and uncertainty issues of the climate data they used. Kotlarski et al 2014, Katragkou et al 2015 in Geosc Mod Dev and Garcia Diez et al, 2015 in Clim Dyn, provide details on the uncertainty issues and biases that stem from different EURO-CORDEX ensemble members. Authors could inform the readers in a concise way about the uncertainty (spread) associated with each meteorological variable used in the statistical model (temperature, precipitation, radiation) and how this may impact the results presented.

The model used to train the statistical model is the CTM CHIMERE forced by the RCM WRF that also contributed to the EuroCordex papers cited by the reviewer under the label "WRF-IPSL-INERIS". We added the following paragraph p 28369, line 3 of the manuscript (page 9, line 7 of the revised manuscript).

> "In the general EuroCordex evaluation, (Kotlarski et al., 2014) finds a good reproduction of the spatial temperature variability even if the models exhibit an underestimation of temperature during the winter in the north Eastern Europe. In addition to this general feature, the specificity of the WRF-IPSL-INERIS member is an overestimation of winter temperatures in the southeast. In terms of precipitations, most of the models exhibit a pronounced wet bias over most subdomains.
> When focusing on WRF members of the EuroCordex ensemble, (Katragkou et al., 2015) points out that the IPSL-INERIS member offers one of the best balance between precipitation and temperature skills."

The manuscript needs language editing (grammar, syntax, expression, punctuation). The title could be changed to become more informative, including more specific keywords such as "statistical model" instead of "proxies", "ozone" instead of "air pollution"

We have modified the title to replace "proxies" by "statistical model". All the technical comments have been taken into account.

Technical comments:

- Better to avoid using 3 dots for not complete lists, better use etc (e.g. page 28364, line 22)
- Abstract. Line 15. Replace "resp." with respectively
- Introduction. Page 28363, line 8/9. "Hence the need to characterize…" the sentence is not complete.
- Page 28363, line 27. "The lack of multi-model approach in air quality and climate projections…" I think it should be in "air quality" or "climate/chemistry" projections and not "air quality and climate" projections, since there is no lack in climate projection ensembles.

- Page 28364, line 19. "have" replace with "has" Page 28364, line 1-4. Please rephrase.

Methodology:

- Page 28365, line 15. What do you mean by "phenomenological"?
  The word phenomenological was referring to chemical and physical processes. In order to avoid misunderstanding, we rephrased the sentence: "The choice of these meteorological variables is based on an analysis of the literature on the chemical and physical processes bearing upon air pollution and meteorology."
- Page 28367, lines 21-23: "By using... over Europe". Please rephrase.
  The rephrased sentence now reads: "By using deterministic climate and chemistry models from the global to the regional scale, they could produce long term air quality projections over Europe".
- Page 28368, lines 24-26: This is a multi-model ensemble consisting of 12 members, 7 out of which are based on the same regional climate model. This implies that climatic information will be too much on the side of the RCA4 climate patterns. Authors could shortly discuss.
  There is indeed an imbalance in the matrix of GCM/RCM in the Eurocordex ensemble. This is an issue for the regional climate modelling community, and we have pointed that out in the revised manuscript page 9 line 17.
- Page 28369, lines 5-7. Could you please be more explicit?
  We rephrased to be more explicit (page 9, line 15): "Both studies are limited to the evaluation of RCM used with perfect boundary conditions (ERA-Interim forcing) and no published study has yet evaluated the various global/regional combinations."
- Page 28370, line 16 (and Figure 1 caption). Authors should add in the text and in the figure caption the model suite out of which results are taken.
  A reference to the model suite has been added.
- Page 28370, line 13. You can skip "The" in "The Figure 1e corresponds to ...".
  This change has been made.
- Page 28371, line 24. You don't define SIA and SOA in text.
  According to the reviewer comment, we have defined Secondary Inorganic Aerosol and Secondary Organic Aerosol.
- Page 28372, lines 8- ..., please explain how you calculate R square/NRMSE for each subregion.
  The following paragraph has been added to the text (page 12, line 7): "The NRMSE is defined as the root mean square error between statistically predicted and deterministically modelled concentrations changes aggregated by region and at daily temporal frequency, normalized by the standard deviation of the deterministic model. It allows describing the predictive power of a model, if the NRMSE is lower or equal to 1 then the model is a better predictor of the data than the mean (Thunis et al., 2012)."
- Page 28373, line 1-2. "This is because... long range transport of air pollution". Can you provide a reference for that? Why couldn't that be an impact of boundaries?
  Long range transport indeed includes pollution advected at hemispheric scale, therefore beyond the European boundaries covered by the model here. However, in the present context, boundary conditions are kept constant as explained section 2.2. Therefore, the long range transport we refer to here, is limited to interconnection within the European domain. It has been rephrased in the text (page 14, line 21).
- Page 28373, line 4-6: Where do we see the lower variability of temperature and incoming SW radiation? Do you suggest that Temperature and SW radiation are not relevant meteorological variables to explain O3 variability in south Europe? This does not make sense to me.
  We are not questioning the impact of temperature and incoming short wave radiation on ozone chemistry. However as a matter of fact they are not identified by the objective statistical analysis as the most discriminating variable because of their more limited range of variation in that region compared to other parts of Europe (standard deviation of 12.5 °C and 150 W/m² for MD; from 15 to 20°C and from 220 to 300 W/m² for the other regions). This has been added page 14, line 26.
- Section 4. This section could be more carefully written; with better structure and proper reference of uncertainty issues available in published literature (see general comments above).
  We have rewritten the section 4, following the general comments of the reviewer.
- Page 28374, line 17. How do you explain the fact that surface temperature increases significantly (did you check for significance? Better to avoid using if not properly tested), and PBL does not change notably? Any references that you could use?

Both the increase of surface temperature and PBL change are statistically significant (Student t-test with Welch variant at the 95% confidence level based on the annual distribution aggregated per region) but the temperature increase is larger than the evolution of the PBL height. This is because surface temperature is not the only the driver of PBL depth (Menut et al., 2013). To address the reviewer comment the following text has been added page 16, line 1):

> "The comparison of historical and future distributions shows that both meteorological drivers evolve significantly in statistical terms (Student t-test with Welch variant at the 95% confidence level based on annual mean). However, even though the PBL depth constitutes the most important meteorological driver for PM2.5, it does not evolve notably compared to the surface temperature in the future (Figure 5Figure)."

- Page 28377, line 5. I don't think you should use the word "slightly".
  The word "slightly" has been removed from the text.

Colette, A., Bessagnet, B., Vautard, R., Szopa, S., Rao, S., Schucht, S., Klimont, Z., Menut, L., Clain, G., Meleux, F., Curci, G., and Rouïl, L.: European atmosphere in 2050, a regional air quality and climate perspective under CMIP5 scenarios, Atmos. Chem. Phys., 13, 7451-7471, 10.5194/acp-13-7451-2013, 2013.

Hedegaard, G. B., Christensen, J. H., and Brandt, J.: The relative importance of impacts from climate change vs. emissions change on air pollution levels in the 21st century, Atmos. Chem. Phys., 13, 3569-3585, 2013.

Jacob, D., Petersen, J., Eggert, B., Alias, A., Christensen, O., Bouwer, L., Braun, A., Colette, A., Déqué, M., Georgievski, G., Georgopoulou, E., Gobiet, A., Menut, L., Nikulin, G., Haensler, A., Hempelmann, N., Jones, C., Keuler, K., Kovats, S., Kröner, N., Kotlarski, S., Kriegsmann, A., Martin, E., van Meijgaard, E., Moseley, C., Pfeifer, S., Preuschmann, S., Radermacher, C., Radtke, K., Rechid, D., Rounsevell, M., Samuelsson, P., Somot, S., Soussana, J.-F., Teichmann, C., Valentini, R., Vautard, R., Weber, B., and Yiou, P.: EURO-CORDEX: new high-resolution climate change projections for European impact research, Regional Environmental Change, 14, 563-578, 10.1007/s10113-013-0499-2, 2014.

Katragkou, E., García-Díez, M., Vautard, R., Sobolowski, S., Zanis, P., Alexandri, G., Cardoso, R. M., Colette, A., Fernandez, J., Gobiet, A., Goergen, K., Karacostas, T., Knist, S., Mayer, S., Soares, P. M. M., Pytharoulis, I., Tegoulias, I., Tsikerdekis, A., and Jacob, D.: Regional climate hindcast simulations within EURO-CORDEX: evaluation of a WRF multi-physics ensemble, Geosci. Model Dev., 8, 603-618, 10.5194/gmd-8-603-2015, 2015.

Kotlarski, S., Keuler, K., Christensen, O. B., Colette, A., Déqué, M., Gobiet, A., Goergen, K., Jacob, D., Lüthi, D., van Meijgaard, E., Nikulin, G., Schär, C., Teichmann, C., Vautard, R., Warrach-Sagi, K., and Wulfmeyer, V.: Regional climate modeling on European scales: a joint standard evaluation of the EURO-CORDEX RCM ensemble, Geosci. Model Dev., 7, 1297-1333, 10.5194/gmd-7-1297-2014, 2014.

Langner, J., Engardt, M., and Andersson, C.: European summer surface ozone 1990-2100, Atmos. Chem. Phys., 12, 10097-10105, 2012a.

Langner, J., Engardt, M., Baklanov, A., Christensen, J. H., Gauss, M., Geels, C., Hedegaard, G. B., Nuterman, R., Simpson, D., Soares, J., Sofiev, M., Wind, P., and Zakey, A.: A multi-model study of impacts of climate change on surface ozone in Europe, Atmos. Chem. Phys., 12, 10423-10440, 10.5194/acp-12-10423-2012, 2012b.

Manders, A. M. M., van Meijgaard, E., Mues, A. C., Kranenburg, R., van Ulft, L. H., and Schaap, M.: The impact of differences in large-scale circulation output from climate models on the regional modeling of ozone and PM, Atmos. Chem. Phys., 12, 9441-9458, 10.5194/acp-12-9441-2012, 2012.

Meleux, F., Solmon, F., and Giorgi, F.: Increase in summer European ozone amounts due to climate change, Atmospheric Environment, 41, 7577-7587, 2007.

Menut, L., Bessagnet, B., Khvorostyanov, D., Beekmann, M., Blond, N., Colette, A., Coll, I.,

Curci, G., Foret, G., Hodzic, A., Mailler, S., Meleux, F., Monge, J. L., Pison, I., Siour, G., Turquety, S., Valari, M., Vautard, R., and Vivanco, M. G.: CHIMERE 2013: a model for regional atmospheric composition modelling, Geosci. Model Dev., 6, 981-1028, 10.5194/gmd-6-981-2013, 2013.
Thunis, P., Pederzoli, A., and Pernigotti, D.: Performance criteria to evaluate air quality modeling applications, Atmospheric Environment, 59, 476-482, http://dx.doi.org/10.1016/j.atmosenv.2012.05.043, 2012.
Thunis, P., Clappier, A., Pisoni, E., and Degraeuwe, B.: Quantification of non-linearities as a function of time averaging in regional air quality modeling applications, Atmospheric Environment, 103, 263-275, http://dx.doi.org/10.1016/j.atmosenv.2014.12.057, 2015.
Watson, L., Lacressonnière, G., Gauss, M., Engardt, M., Andersson, C., Josse, B., Marécal, V., Nyiri, A., Sobolowski, S., Siour, G., and Vautard, R.: The impact of meteorological forcings on gas phase air pollutants over Europe, Atmospheric Environment, 119, 240-257, http://dx.doi.org/10.1016/j.atmosenv.2015.07.037, 2015.