This paper conducted a multi-model ensemble of regional models to simulate air quality over Europe and the U.S. for 2010 in the frame of AQMEII3. It estimated the impact of air pollution ($PM_{2.5}$, $O_3$, CO and $SO_2$) on human health and the associated external costs over the two continents using a common health assessment approach. Furthermore, the authors also conducted several emission perturbation scenarios to investigate the domestic and foreign contributions to the related health impacts. Overall this study is interesting and the manuscript is relatively logically organized. However, several parts of the manuscript need to be clarified and figures should be plotted more clearly. Following general and specific comments should be addressed before the publication in Atmospheric Chemistry and Physics.

**General comments:**

- The multi-model ensemble approach is widely used, especially in forecast studies in which observations are not available to evaluate the performance of individual models. Here the authors use multi-model ensemble results to investigate the air pollution levels in 2010, where sufficient measurements are available over Europe and the U.S. Therefore, the authors should show that the ensemble results are better than any individual models. As shown in Table 3 and Table 6, the RSME of multi-model ensemble results ($MM_m$ and $MM_{opt}$) are even larger than those of individual model results. Since the equations and datasets used to calculate these statistics in Tables 3 and 6 are unclear, it is difficult to judge the performance of the ensemble results. Particularly, the DE1_SMOKE simulation over the U.S. significantly underestimates $SO_2$, CO, and $PM_{2.5}$ (even up to a factor of three) comparing with the observations, which means that this result has systematic bias. This model should be removed from the ensemble, but I am not sure how it is being treated in the optimal-reduced multi-model ensembles. More description and explanations are needed here.

-This study mainly focuses on estimating the air pollution related health impacts, where annual mean concentrations of CO, $SO_2$ and $PM_{2.5}$ and yearly sum of daily maximum 8-hour $O_3$ running average over 35 ppb are used in the EVA system. The model evaluation in Section 3.1 should focus more on the spatial distribution of these models' performance, rather than on the average over the whole region. Furthermore, the authors should provide more necessary information for model evaluation, e.g., sources of observations, equations used to calculate the statistics, etc.

-From the model evaluation, it shows that results from different models have large divergence. This should be caused by many factors, like emissions, transport, chemistry, dry/wet removals. I would suggest the authors provide more information about the mechanisms/parameterizations used for each model in the supporting materials.

-In this study, the intercontinental impacts are investigated using the 20 % emission reduction scenarios applied over the source regions. In their model experiments, a

global model was used to provide chemical boundary conditions for all participating regional models. To my knowledge, the long-range transport of air pollutants is controlled by many complicated factors, which may lead to much larger uncertainties over the long-distance path than the source region. I am not sure that using a single model to represent the long-range transport is a proper way for an ensemble analysis. Therefore, the authors should provide more information regarding the evaluation of the global model.

- Figure quality is low and needs improvement, especially for Figures 1 and 4. The authors should make font-size, colorbar size, subtitles, units, and plot captions consistent. See specific comments below.

**Specific comments:**

Lines 102-116: This paragraph introduced a number of previous works quantifying air pollution-related health impacts due to intercontinental transport. However, the results of those studies showed inconsistent relative importance of domestic versus foreign emissions. Please comment on this.

Lines 250-251: " … previous AQMEII-related works" need to show some references here.

Lines 254-255: The authors should briefly introduce the sources and features of these observation data used in this study.

Lines 329-330: The authors should describe in detail how the observed and simulated monthly time series in Figures 2 and 3 are obtained. For example, whether or not the observed and simulated results averaged over the whole continental regions are sampled with identical time and locations.

Lines 390-391: "…the numbers of cases are strongly correlated to the population density…", please refers to Figure 1 for comparison.

Table 6: Why not use the same units for Europe and North America?

Figure 1: Please clarify which continent the left/right panel refers to in the caption. The unit of population density also needs to be provided. More detailed terrestrial boundaries are recommended to distinguish countries or states. Furthermore, I recommend using the same scale for the two panels to have a better comparison.

Figure 4: besides the same comments for Figure 1, figure quality needs to be improved significantly. The authors should be consistent in making the plots. For example, the top two plots have subtitles while the bottom ones don't. The font-size and colorbar

size of these panels are different. The units are missed in the top two panels. The colorbar of plot (d) even overlaps the coordinate. Additionally, the caption does not provide all necessary information to understand this figure.