We thank the reviewer for the review. We have tried to implement all the comments and corrections in the new manuscript.

General comments:

Comment: 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 (MMm and MMopt) 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 SO2, CO, and PM2.5 (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.

Response: We have now extended the description and the discussion on mean and median multi-model results (Lines 482-484; 511-515; 547-550; 611-614). In order to be able to estimate an uncertainty in the health impacts calculations using concentration inputs from different models, none of the models were removed from the ensemble. It is true that the multi model mean results do not outscore all individual models and that is why we present both individual model results and multi-model ensemble results in the manuscript.

Comment: This study mainly focuses on estimating the air pollution related health impacts, where annual mean concentrations of CO, SO2 and PM2.5 and yearly sum of daily maximum 8-hour O3 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.

Response: We have now added spatial model performance based on the bias (Figures 4 and 5) and included the relevant discussion (Lines 485-499; 516-528).

Comment: 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.

Response: We have now added more details in Table 1 and model system descriptions in the supplementary material adopted from Solazzo et al., 2017.

Comment: 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.

Response: We have used one global model to produce the boundary conditions to the regional CTMs in order to limit the uncertainty in the multi-model ensemble. The evaluation of the global is not the aim of this study s it is a common input to all the regional models. C-IFS model has been extensively evaluated elsewhere (e.g. Flemming et al. (2015 and 2017), and in particular for the North America in Hogrefe et al. (2017) and Huang et al. (2017).

Comment: 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.

Response: We have now improved the figures.

Specific comments:

Comment: 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.

Response: These studies uses different sets of global models on different spatial resolutions. However results were consistent in terms of the contribution of local vs. non-local sources on the impacts of pollution.

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

Response: These references are already listed in Lines 301-302.

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

Response: We have now added information on the source of the observations (Lines 250-259): “The observational data used in this study are the same as the dataset used in second phase of AQMEII (Im et al., 2015a, b). Surface observations are provided in the Ensmeble system (http://ensemble2.jrc.ec.europa.eu/public/) that is hosted at the Joint Research Centre (JRC). Observational data were originally derived from the surface air quality monitoring networks operating in EU and NA. In EU, surface data were provided by the European Monitoring and Evaluation Programme (EMEP, 2003; http://www.emep.int/) and the European Air Quality Database (AirBase; http://acm.eionet.europa.eu/databases/airbase/). In NA observational data were obtained from the NAtChem (Canadian National Atmospheric Chemistry) database and from the Analysis Facility operated by Environment Canada (http://www.ec.gc.ca/natchem/).”

Comment: 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.

Response: We have now added the following (Lines 244-250): “The models’ performance on simulating the surface concentrations of the health-related pollutants were evaluated using Pearson’s Correlation (*r*), normalized mean bias (*NMB*), normalized mean gross error (*NMGE*) and root mean square error (*RMSE*) to compare the modelled and observed hourly pollutant concentrations over surface measurement stations in the simulation domains. The hourly modelled vs. observed pairs are averaged and compared on a monthly basis. The modelled hourly concentrations were first filtered based on observation availability before the averaging has been performed.”

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

Response: We have now referred to Fig. 1 (Line 590).

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

Response: We have now corrected the captions. Units are consistent over the two domains.

Comment: 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.

Response We have now updated Fig. 1.

Comment: 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.

Response: We have now updated Fig. 4 (now Fig. 6).