We thank the reviewer for the comments and corrections. We have now implemented all the points to further develop the manuscript.

General comments

Comment: First, the description of the health impact assessments and the economic impacts should be more detailed, and include especially all the assumptions and choices made in making the computations and assessments. There are numerous alternative choices that you will need to make for e.g. economic evaluations; some of these have been properly described and discussed, whereas some have not been described. Reviewer number 1 has already detailed this issue.

Response: The EVA methodology section has been substantially extended (Lines 326-464).

Comment: Second, there are also gaps in the description of the individual CTM's and, the constructed ensemble and the evaluation of the models and the ensemble. In particular, there is very little discussion on how the non-anthropogenic emission sources have been included; as these constitute a substantial part of the total PM mass, these should also be described. There should be also discussion on the main limitations of the CTM's and the emission inventories used, what are their main uncertainties and the most poorly known parts of modelling. Details on this issue are in 'detailed comments'.

Response: We have added more details in Table 1 and added model descriptions to the supplementary materials adopted from Solazzo et al. (2017).

Comment: Regarding model evaluation, the manuscript should specify which networks of stations were used, how many stations were considered within each domain, and what were their site classifications. Large PM deficits were found for some models. The manuscript should therefore discuss the most probably reasons for these underpredictions: were these caused by deficiencies of the used CTM's, missing emissions or both, or/and some other reason.

Response: We have extended the model evaluation part (Lines 485-499; 516-528).

Comment: Regarding the presentation of the results, there are a lot of large tables, but in my view too little synthesis and graphical illustration of the main results and findings. I would recommend to move some of the large tables an annex or to supplementary materials for better readability, and some summary figures could be added instead, to highlight the main insights, findings and conclusions.

Response: We have moved some of the tables (Table 3 and Table 4) in the supplement and kept the ensemble mean results together with the optimal ensemble results from old Table 7 to the new Table 3. However, we believe that these numbers should be explicitly presented in the manuscript

as particularly the morbidity calculations are for the first time calculated for both continents and transferring them into figures would lose the details.

Comment: Regarding the section 'materials and methods', I recommend to use the traditional sections for a better readability, e.g., first Evaluation of emissions, then Atmospheric dispersion modelling, the construction of ensembles, Health impact assessment and finally economic parts. The current subtitles list one project and one model.

Response: We have now re-structured this section following the reviewers recommendations.

Detailed comments

Abstract.

Comment: Lines. 52-53. This is one of the main results of the study, so it should be presented clearly. This study addresses models for (i) emissions, (ii) dispersion, (iii) health assessment and (iv) economic evaluation. The term 'model' should therefore be used carefully and specified as necessary, throughout the manuscript. This sentence probably refers to CTM's but not health models (or emission models). It is therefore variation due to the differences of CTM's. However, the computed health impacts can also vary a lot depending on which health assessment model would be used, and which health assessment assumptions would be selected. In this study, the authors have addressed the variability due to CTM's but not that of the health assessment modelling, although the latter uncertainty is commonly much larger. Please clarify and write more clearly and accurately what is meant.

Response: We have now rephrased this sentence accordingly (Lines 53-55): "Health impacts estimated by using concentration inputs from different chemistry and transport models (CTMs) to the EVA system can vary up to a factor of three in Europe (twelve models) and the United States (three models)."

Comment: Lines 54-55. These results could be also presented per capita; this would better illustrate better the differences of the two selected domains. The PM concentration levels and the distributions of population of the two domains could also be quantitatively compared. 'In agreement', specify quantitatively, e.g., within what percentage.

Response: We have now added normalized PD numbers (number deaths per 100 000) in the text.

Comment: Line 68. Write the acronym in full.

Response: We have provided the full name of the acronym (Lines 48-52): "Along with a base case simulation, additional runs were performed introducing 20% anthropogenic emission reductions

both globally and regionally in Europe, North America and East Asia, as defined by the second phase of the Task Force on Hemispheric Transport of Air Pollution (TF-HTAP2)."

Comment: Line 71. 'global anthropogenic emissions' – specified for which pollutant species ?

Response: Emission perturbations target anthropogenic emissions. This is now made clear in the text (Lines 63-71): "A total of 54 000 and 27 500 premature deaths can be avoided by a 20% reduction of global anthropogenic emissions in Europe and the U.S., respectively. A 20% reduction of North American anthropogenic emissions avoids a total premature death of ~1 000 in Europe and 25 000 total premature deaths in the U.S. A 20% decrease of anthropogenic emissions within the European source region avoids a total premature death of 47 000 in Europe. Reducing the East Asian anthropogenic emissions by 20% avoids ~2000 total premature deaths in the U.S. These results show that the domestic emissions make the largest impacts on premature death, while foreign sources make a minor contributing to adverse impacts of air pollution."

Comment: Line 72. 'emissions foreign emission' – correct sentence

Response: The sentence has been corrected (Lines 64-66).

Comment: Lines 75-77. 'foreign sources make a minor contributing : : :'. This is too general. Whether the sources in a specified domain contribute more or less to health within that domain depends on a lot of factors, such as e.g., population densities in the considered areas, how large the considered two areas are, which pollutants are considered, etc. This statement is therefore correct for some cases, and not correct for some others. Please rewrite the statement more accurately.

Response: We agree with the reviewer. However, the abstract is just an overall short summary of the paper so such a discussion does not fit to this section. We have now slightly rephrased the sentence as following: "These results show that the domestic anthropogenic emissions make the largest impacts on premature death on a continental scale, while foreign sources make a minor contributing to adverse impacts of air pollution."

Introduction

Comment: Lines 107-109, and lines 114-117. Same comment as above. Whether these statements are true, depends on various factors – the relevant factors therefore need to be specified.

Response: These studies employ global model ensembles on coarse spatial resolutions to calculate mortality due to air pollution.

Comment: Lines 134-136. When presenting cost values, it is proper to state also for which year this has been evaluated.

Response: The currency year is 2013 (Lines 463-464).

Comment: Line 168. ': : : seen : : : ' - correct the English language.

Response:. We have rephrased the sentence as following (Lines 171-173): "Source-receptor relationships have the advantage of reducing the computing time significantly and have therefore been extensively used in systems like GAINS (Amann et al., 2011)."

Comment: Lines 200-202. Using a so-called optimal ensemble is fine, but as far as I know, it does not guarantee that there is e.g. no redundancy or recursiveness of models. Practically in all cases, a collection of CTM's will have some very similar treatments; using an 'optimal' ensemble will probably reduce their effect, and that is OK, but it does not altogether remove these effects.

Response: We agree with the reviewer. That is why we write that we produce an optimal ensemble producing the minimum error at each time step for each pollutant, and do not say that we remove the error altogether.

Materials and methods

Comment: Line 218. Should read 'emission information'. There are also several other input datasets, obviously. Report also the modelling of sea salt, desert dust, biogenic emissions, wild-land fires, etc. Add some discussion on what were the main limitations, uncertainties and gaps of modelling of the CTM's used.

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

Results

Comment: What were the networks of stations used in Europe and the US; these should be described. How many stations were considered ? What were the classifications of stations – were all of these classified as regional or global background ?

Response: We have extended the model evaluation section (Lines 244-263).

Conclusions

Comment: Line 562. This statement may be true, but it should be supported by quantitative evidence: were there model runs to quantify this effect, and how large was it in e.g. per cents of predicted concentrations ? Alternatively, if not confirmed, this statement could be removed.

Response: This is the most important gap in air pollution-related health studies and therefore needs to be investigated. Therefore, there are no studies yet that designed such an experiment. Further down, we refer to a Nordic project that works on these issues.

Comment: Lines 533-538. The underestimation of PM mass is a key uncertainty. There should therefore be some accurate assessment on the reasons resulting to this uncertainty. For instance, 'natural emissions' are mentioned, but it is not stated in the text which of these were included, which were neglected, and which possible omission or underestimation could probably have the largest effect. Please add some discussion of the most probable causes of the under-prediction.

Response: We have now extended this paragraph (Lines 748-754). As shown in the supplementary material, the CTMs diverge a lot on the representation of particles and their size distribution, SOA formation, as well as the inclusion of natural sources. As the anthropogenic emissions are harmonized in the models, they represent a minor uncertainty in terms of model-to-model variation. However, differences in the treatment of the temporal, vertical and chemical distributions of the particulate and volatile organic species have an influence in the model calculations and therefore lead to model-to-model variations.