Response to Anonymous Referee #1

We thank the reviewer for the constructive and positive overall assessment of our paper. In the following, we respond to the individual points raised in italic letters, the reviewer comments are left in normal font.

This paper presents a summary of the works that were conducted within the frame of the European ECLIPSE project. First the overall strategy of the project is declined and then, the results obtained at every step of the project are exposed and discussed. The choice of a multi-angle approach (SRF, GTP, model scenarios) brings strength to the results and allows providing important and new advances for the climate change research community. The paper is well written and organized and the quite large number of results is presented in a clear and concise way. Furthermore, the discussion about the consistency of the 2 paths of research is appreciable and allows the identification of future needs for climate research studies. As this article is a presentation paper, the methods and results of the works that exposed in the accompanying articles will not be commented, and I will rather focus on the way there are highlighted and discussed.

Thank you for these comments.

My comments and questions are the following: 1) The comparisons between the model outputs and the observations show an insufficient degree of restitution of the gas & particle concentrations in different regions of the Earth atmosphere. However, this lack of representativeness of the models is only mentioned, but it is not considered in the discussion in the rest of the paper. Ex : Does the amplitude of BC underestimation in the Arctic or the overestimation of SO_2 in the continental atmosphere strongly affect the final result about the impact of BC and SO_2 emission reductions on the final temperature increase? The compensation between the impact of OA and BC may also be altered by a wrong representation of the BC to OA ratios in the models. Furthermore, it is not mentioned how/if the improvement of OA reactivity and formation (or BC lifetime changes) in the ECLIPSE models quantitatively changed the model predictions compared with observations. It is not fully satisfactory to consider that the models well restitute the NO_x concentrations when they both under- and over- estimate observations. The issue of NO_x is a problem of proximity and nothing is said about the regionalization of over- and under- estimation. And the evaluation of the efficiency of a scenario (on an air quality basis) relies on the exposure of citizens to high NO_x concentrations. Finally, the reader there may wonder if the differences between the results of the models in the reference run are linked with the differences in the predicted impacts of emission control on climate change. Then, the predictions of a given model that under- or over- estimates the concentrations of a given SLCP could be interpreted in the light of its comparison to measurements.

This comment touches upon many different aspects of our paper. We find it somewhat difficult to address, as this would require addition of substantial detail and discussion which we deem unsuitable for an overview paper such as this, especially given that it is already very long. Most answers to the above questions are also different for every model and adding all this information would distract from the concept of an ensemble-average view. We refer to the specialized papers for this (e.g., Tsigaridis et al., 2014, for the OA issue; or Samset et al., 2014, for the BC lifetime) and have tried to make these references more explicit and clear. In addition, we have tried to add qualifying notes at several places. For instance, with respect to *the importance of BC errors in the Arctic, we have added in section 3.2 (before the actual comparisons):*

"Aerosol loadings in the Arctic are generally much lower than in populated regions and the Arctic encompasses only a small fraction of the Earth. Therefore, impacts of even large relative errors in the modelled aerosol concentrations in the Arctic on global radiative forcing and global climate response are relatively small. Nevertheless, identification of model biases in this remote region is important as it can lead to improved process understanding, especially of the aerosol removal mechanisms."

On the issue of NOx, for instance, we have added: "For rural NO_2, also the individual models deviate less from the measured concentrations than for the urban stations, indicating that the individual model biases for urban NO_2 are very likely mainly due to the limited model resolution and not to biases in emissions and/or chemical processes."

We have also added a reference to one publically available ECLIPSE deliverable report (Schulz et al., 2015), which addresses several of the points raised by the reviewer, e.g., the SO2 and NOX emission issue.

2) The same questions arise concerning the ability of the models to restitute past behaviors. Figure 6 indeed shows that models reproduce the past changes in the warming trend during the 1990-2005 period, but the absolute amplitude of the trend is not well captured by the models. Despite this, these discrepancies are not considered as a limitation for the interpretation of model predictions in the rest of the discussion.

Indeed, the magnitude of the trends is not so well captured by all models. However, it is virtually impossible to correct for errors because 1) the comparison was made only for a small region and it is not well known how the models have captured trends in other regions; 2) the observational database even for the study region is limited; 3) natural climate variability contributes to the trends; 4) the models used for the climate response simulations have been improved in terms of emissions and processes. Therefore, the best use of this comparison is to point out the limitations of current model capabilities, as suggested by the reviewer. To make this more clear, we have added the statement: "However, the absolute amplitude of the trends is not equally well captured by all models, indicating that the skill of the ECLIPSE (and other) ESMs to simulate temperature trends responding to changing aerosol emissions is limited. This is due to both limitations in the models themselves, the emission input, as well as the influence of natural climate variability."

3) Concerning the models, insufficient information is given about their differences and similarities, which is a crucial point when running an ensemble. In particular, when looking at the diversity of responses to BC forcing, the way they consider the direct and semidirect effects of BC should be detailed. Secondly, the constitution of a model ensemble is also questioning, due to the low number of models running on a same compound for some experiments. The differences and similarities in the model structures thus strongly impact the amplitude of the answers. Considering this, the mean model response as well as the range of their responses to emission perturbations have to be interpreted with caution.

Indeed the number of models is relatively small. However, for any European project it would have been difficult to gather more models than were involved in ECLIPSE. Larger ensembles require global efforts but these are usually less focussed on specific research questions. The four ECLIPSE Earth System models were also among the 15 global models contributing to the AeroCom direct aerosol forcing estimate of BC (Myhre et al., 2013). Although, the mean *of the four Eclipse models was somewhat stronger than the mean of the 15 AeroCom models, the four Eclipse models span the range of the AeroCom models from the strongest estimate to almost the lowest estimate. Therefore, while the number of models was small, we consider them largely representative of the current generation of ESMs. Nevertheless, we have added as a final conclusion of our paper:*

"The number of models contributing to the ECLIPSE project was relatively small. While the models were shown to be largely representative of results obtained from larger model ensembles, this makes quantification of mean values and especially uncertainties (e.g., of RF or temperature response) dependent on the particular properties of the ECLIPSE models. For a more comprehensive quantification of uncertainties, it is therefore recommended to repeat the modelling exercises presented in this paper with a larger international model ensemble."

The reviewer also asks for more information about the models. However, none of the models involved in ECLIPSE is new and all of them have been described extensively in the literature. In Table 2, we give references to the most recent descriptions of all the models involved. Furthermore, several specialized ECLIPSE papers have described the models and their properties in more detail (e.g., Baker et al., 2015). We therefore think that it is neither necessary nor commendable to overload this overview paper with extensive details about individual models.

4) About the constitution of the MIT scenario: the mitigation basket is obtained through a selection of emission control measures on the basis of their potential for reducing the global warming. Such a procedure asks several questions: On which basis are the options combined? Are the set of measures consistent in terms of operational set-up? Isn't there a possibility that the combination of several measures is not politically or financially realistic? Finally, is the basket realistic for Air Quality ? Indeed, as the selection is based on GTP20, it is possibly not the most expectable basket for air quality. There may be other actions (considered as more efficient) that will have to be considered in the future years to reduce the exposure of urban citizens to air pollution, and it would have been interesting to consider their potential for limiting the global warming rather than considering only SLCP actions on the basis of their GTP20. Of course, rethinking the MIT scenario is not the purpose of this presentation paper, but the discussion about the fact that "the co-benefits of the non-CH4 SLCP mitigation measures are quite limited" is strongly affected by the constitution of the mitigation basket, and this may be highlighted. This is important because the evaluation of the Air Quality Impacts of Short-Lived Pollutants is one focus of the paper, as mentioned in the title.

The MIT scenario was designed following one of the key objectives of the ECLIPSE project: the evaluation of climate impacts of SLCP mitigation and air quality co-benefits of such a strategy. The health aspects of such strategies are not ignored but were not the primary objective for developing the scenario. MIT combines all measures yielding a beneficial climate impact in terms of GTP_{20} while also improving air quality. The measures have all been derived with the GAINS model and are internally consistent and realistic in a sense of considering their simultaneous impacts on emissions of all key co-emitted species. Of course, our scenario is unlikely to ever be implemented exactly as such, as real-world scenarios will consider a number of local factors and priorities, including consideration of costs of the measures which might limit or delay implementation of proposed measures. This was not considered by the ECLIPSE project. We have also not considered any actions that would be beneficial for air quality but which would have a detrimental effect on climate (in terms of GTP_{20}), or vice versa. Generally, a package of measures optimized for improving air quality would include several control technologies reducing sulphur dioxide emissions rendering a less beneficial climate impact than the basket considered here unless additional (and strong) CO_2 mitigation would be included. Such a scenario is indeed of high relevance and interest for the community but it has not been developed and analysed within the ECLIPSE project. The fact that we considered only one scenario was entirely due to the resource limitations prescribed by the transient climate runs. It was *simply not possible to run ensemble simulations for several such scenarios. We consider the MIT scenario as the best choice given these limitations.*

On the choice of the mitigation basket, we have also added the following bullet point in the conclusion section:
"The ECLIPSE mitigation scenario has been developed to be representative of a mitigation strategy that considers both climate and air quality, assuring reduced climate forcing without detrimental impact on air quality; as a matter of fact, strong air quality co-benefits were identified. Real-world scenarios are likely to favour particular policy objectives and will also consider the costs for the mitigation measures. More work is therefore needed to explore a larger range of scenarios. By demonstrating the efficiency and capacity of the metrics-based approach to quantify temperature changes and its consistency with transient climate model simulations, ECLIPSE has opened the way to explore a large number of such scenarios. This would be an impossible task if transient climate ensemble model simulations were needed for each."

More technically: The gain in air quality brought by the MIT scenario compared with the CLE scenario is shown for ozone and PM2.5, but the absolute improvement in the concentrations, as modelled in the CLE scenario, is not shown (it is just rapidly mentioned in the text). It makes difficult the appreciation of the gain of the mitigation scenario.

Table 4 lists the difference in surface concentrations for ozone and $PM_{2.5}$ between the MIT and CLE scenario for various regions of the world. While, indeed, this is a concise summary, we believe nevertheless that this clearly (and quantitatively) reports the improvements due to the mitigation.

Additional references:

Myhre, G., Samset, B. H., Schulz, M., Balkanski, Y., Bauer, S., Berntsen, T. K., Bian, H., Bellouin, N., Chin, M., Diehl, T., Easter, R. C., Feichter, J., Ghan, S. J., Hauglustaine, D., Iversen, T., Kinne, S., Kirkevag, A., Lamarque, J. F., Lin, G., Liu, X., Lund, M. T., Luo, G., Ma, X., van Noije, T., Penner, J. E., Rasch, P. J., Ruiz, A., Seland, O., Skeie, R. B., Stier, P., Takemura, T., Tsigaridis, K., Wang, P., Wang, Z., Xu, L., Yu, H., Yu, F., Yoon, J. H., Zhang, K., Zhang, H. and Zhou, C.: Radiative forcing of the direct aerosol effect from AeroCom Phase II simulations, Atmos. Chem. Phys., 13, 1853-1877, 2013.

Schulz, M., Olivié, D., Tsyro, S., Kanakidou, M., Myriokefalitakis, S., Daskalakis, N., Im, U., Fanourgakis, G., Hodnegrog, Ø., Skeie, R., Lund, M., Myhre, G., Bellouin, N., Rumbold, S., Collins, B., Cherian, R., and Quaas, J.: ECLIPSE Deliverable 2.1: Report on model accuracy, available from http://eclipse.nilu.no/language/en-GB/ProjectOverview/Deliverables.aspx.