REPLIES TO REVIEWER #1

We thank reviewer 1 for the insightful comments, and for pointing to inconsistencies. We apologize for needing more time than anticipated to address all comments, but we believe that we have been able to address most issues, and that we have significantly strengthened the manuscript.

Before addressing the comments we would like to mention that we have modified the abbreviation of the O_3 health exposure metric (6 monthly daily maximum 1-h concentration) from M6M to 6mDMA1 (and accordingly M3M to 3mDMA1) as the latter seems to be commonly used in other works.

In the following we have placed the numbered reviewer comments in boxes. Our reply to the reviewer is in blue font, the changes to the manuscript in red font.

We also attach a revised version of manuscript and supplement with tracked changes compared to the first version.

REVIEWER 1 comments:

The manuscript presents a detailed summary of the methodology and validation for the TM5-FASST screening tool. TM5-FASST is a simplified tool that uses linear source receptor relationships of air pollutant precursor species across 56 geographical source regions (plus aviation and shipping) to calculate the response in air pollutant concentrations at both the surface and 25 vertical layers in the atmosphere. The difference in concentrations can then be used to calculate the change in a number of air pollution impact metrics related to human health, climate and crop production. The tool allows for the impact from different emissions scenarios to be explored without the need to run more detailed composition climate models. The manuscript provides a through description of the underlying methodology of TM5-FASST as well as an evaluation of the air pollutant predictions and impact metrics against a number of different sources. It provides a good reference for the TM5-FASST tool for use in future studies.

Major General Comments

1) Whilst I understand that TM5-FASST is not meant to replicate full scale model simulations, it would be good to bring together the limitations together into a more coherent section, possibly within the discussion section. Throughout the manuscript specific sections of the text mention aspects that TM5-FASST will not be able to predict e.g. changing spatial distribution of emissions and chemical regime. It would make sense for the reader to have these all in one place.

REPLY: Thank you for the suggestion. We have substantially edited and extended the discussion section to address the limitations of the tool..

CHANGES TO MANUSCRIPT: *New section 4*:

4 Discussion

Although the methodology of a reduced-form air quality model, based on linearized emission – concentration sensitivities is not new and has been successfully applied in earlier studies (Alcamo et al., 1990), the concept of directly linking pollutant emission scenarios to a large set of impacts across various policy fields, in a global framework, have made TM5-FASST a highly requested tool in a broad field of applications. HTAP1 showed that TM5 source-receptor results

(for the large HTAP1 regions) were in most cases similar to the median model results of more than 10 global models, lending additional trust to the model performance (e.g. Anenberg et al., 2014; Dentener et al., 2010; Fiore et al., 2009). The results in the previous sections have outlined its strengths and weaknesses. The major strength of the tool is its mathematical simplicity allowing for a quick processing of large sets of scenarios or scenario ensembles. An extreme example is the full family of SSP scenarios delivered by all participating Integrated Assessment Models, for decadal time slices up to 2050, constituting a batch of 594 scenarios of which a selection of 124 scenarios was analysed with TM5-FASST in the study by Rao et al. (2017). Further, the tool is unique in having a broad portfolio of implemented impact modules which are evaluated consistently over the global domain from the same underlying pollutant field which creates a basis for a balanced evaluation of trade-offs and benefits attached to policy options.
On the other hand, the reduced-form approach inevitably encompasses a number of caveats and uncertainties that have to be considered with care and which are discussed in the following sections.

4.1 Issues related to the reduced-form approach

The reliability of the model output in terms of impacts depends critically on the validity of the linearity assumption for the relevant exposure metrics (in particular secondary components), which becomes an issue when evaluating emission scenarios that deviate strongly from the base and -20% perturbation on which the current FASST SRs are based. The evaluation exercise indicated that non-linearity effects in $PM_{2.5}$ and O_3 metrics in general lead to a higher bias for stringent emission reductions (towards -80% and beyond) than for strong emission increases compared to the RCP2000 base case, but over-all remain within acceptable limits when considering impacts. Indeed, because of the thresholds included in exposure-response functions, the higher uncertainty on low (below-threshold) pollutant levels from strong emission reductions has a low weight in the quantification of most impacts. In future developments the available extended-range (-80%, +100%) emission perturbation simulations could form the basis of a more sophisticated parameterization including a bias correction based on second order terms following the approach by Wild et al. (2012) both for O_3 and secondary $PM_{2.5}$. The break-down of the linearity at low emission strengths is relevant for O_3 and O_3 exposure metrics as the implementation of control measures in Europe and the US has already substantially lowered NO_x levels over the past decade, gradually modifying the prevailing O_3 formation regime from NO_x-saturated (titration regime) to NO_x-limited (Jin et al., 2017).
Ozone impact on agricultural crop production is deemed to be the least robustly quantified impact category included in FASST, in particular when evaluated from the threshold-based AOT40 metric, and has to be interpreted as indicative order-of-magnitude estimate. In an integrated assessment perspective of evaluating trade-offs and benefits of air pollutants scenarios, the dominant impact category however appears to be human health (Kitous et al., 2017; OECD, 2016; UNEP, 2011) where TM5-FASST provides reliable estimates.
Another issue for caution relates to the FASST analysis of emission scenarios with spatial distribution that differs from the FASST reference scenario (RCP year 2000). The definition of the source regions when establishing the SR matrices implicitly freezes the spatial distribution of pollutant emissions within each region, and therefore the reduced-form model cannot deal with intra-regional spatial shifts in emissions. In practice this is not expected to introduce large errors as anthropogenic emissions are closely linked to populated areas and road networks of which the extent may change, but much less so the spatial distribution. It can be a problem when going far back in time, when large patterns of migration and land development occurred, while in RCP scenarios relatively simple expansions of emissions into the future did not assume huge shifts in regional emission patterns.
The implicitly fixed emission spatial distribution may also become relevant when making a sector apportionment of pollutant concentrations and impacts. Source-Receptor relations are indeed particularly useful to evaluate the apportionment of emission sources (in terms of economic sector as well as source regions) to pollutant levels in a given receptor. However, as

the TM5-FASST_v0 source-receptor matrices were not segregated according to economic sectors, an emission reduction of 20% for a given source region is implicitly considered as a 20% reduction in all sectors simultaneously. Although the atmospheric chemistry and transport of emissions is in principle independent of the specific source, a difference in the sector-specific SR matrices may occur due to differences in temporal and spatial (horizontal/vertical) distribution of the sources. Therefore apportionment studies on sectors which have a significantly different emission spatial distribution than other sectors in the same region should be interpreted with care. In particular impacts of off-shore flaring cannot be assessed with TM5-FASST because those emissions were not included in the RCP base emissions. This limitation however does not apply to international shipping and aviation for which specific SR matrices have been established.

Comparing to earlier studies and reference data, the performance of TM5-FASST with respect to climate metrics is satisfactory, with the exception of BC forcing which is at the low side of current best estimates. In fact, earlier TM5-FASST assessments where climate metrics were provided (UNEP, 2011; UNEP and CCAC, 2016) applied a uniform adjustment factor of 3.6 on BC forcing, in line with the observation by that many models underestimate atmospheric absorption attributable to BC with a factor of almost 3. In TM5-FASST, an adjustment factor of 3.6 leads to a global forcing by anthropogenic BC of 600 mW m^{-2}. This tuning factor implicitly accounts for not-considered BC forcing contributions and for a longer BC atmospheric lifetime than implemented in the TM5 model and the resulting FASST SR coefficients.

The current version of TM5-FASST is missing some source-receptor relations which may introduce a bias in estimated $PM_{2.5}$ and O_3 responses upon emission changes. The omission of secondary organic PM in TM5 is estimated to introduce a low bias in the base concentration of the order of 0.1 µg m^{-3} as global mean however with regional levels in Central Europe and China up to 1 µg m^{-3} in areas where levels of primary organic matter are reaching 20 µg m^{-3} (Farina et al., 2010) indicating a relatively low contribution of SOA to total PM2.5. O_3 formation from CO is included in the TM5 base simulations, but no SR matrices for the FASST source region definition are available. Based on the HTAP1 CO perturbation simulations with TM5, we estimate that a doubling of anthropogenic CO emissions contributes with 1 – 1.9 ppb in annual mean O_3 over Europe, 1.3 -1.9 ppb over North-America, 0.7-1.0 ppb over South Asia and 0.3 – 1.5 ppb over East-Asia. Development of CO-O_3 SRs is an important issue for the further development of the tool.

4.2 Inter-annual meteorological variability

A justified critique on the methodology applied to construct the FASST SRs relates to the use of a single and fixed meteorological year 2001, implying possible unspecified biases in pollutant concentrations and source-receptor matrices compared to using a 'typical meteorological/climatological year'. We followed the choice of the meteorological year 2001 made for the HTAP1 exercise. As the North-Atlantic Oscillation (NAO) is an important mode of the inter-annual variability in pollutant concentrations and long range transport (Christoudias et al., 2012; Li et al., 2002; Pausata et al., 2013; Pope et al., 2018), the HTAP1 expectation was that this year was not an exceptional year for long-rang pollutant transport - e.g. for the North-Atlantic region, as indicated by a North Atlantic Oscillation (NAO) index close to zero for that year (https://www.ncdc.noaa.gov/teleconnections/nao/). The HTAP1 report (Dentener et al., 2010) also suggested that "Inter-annual differences in SR relationships for surface O_3 due to year-to-year meteorological variations are small when evaluated over continental-scale regions. However, these differences may be greater when considering smaller receptor regions or when variations in natural emissions are accounted for". The role of spatial and temporal meteorological variability can thus be reduced by aggregating resulting pollutant levels and impacts as regional and annual averages or aggregates, the approach taken in TM5-FASST.

The impact of the choice of this specific year on the TM5-FASST model uncertainty or possible biases in base concentrations and SR coefficients is not easily quantified. For what concerns the pollutant base concentrations, some insights in the possible relevance of meteorological variability can found in the literature. For example, Anderson et al., (2007) showed that in

Europe, the meteorological component in regional inter-annual variability of pollutant concentrations ranges between 3% and 11% for airborne pollutants (O_3, $PM_{2.5}$), and up to 20% for wet deposition. On a global scale, Liu et al. (2007) demonstrated that the inter-annual variability in PM concentrations, related to inter-annual meteorological variability can even be up to a factor of 3 in the tropics (e.g. over Indonesia) and in the storm track regions. A sample analysis (documented in section S2.2 of the SI) of the RCP year 2000 emission scenario with TM5 at 6°x4° resolution of 5 consecutive meteorological years 2001 to 2005 indicates a year-to-year variability on regional $PM_{2.5}$ within 10% (relative standard deviation) and within 3% for annual mean O_3. We find a similar variability on the magnitudes of 20% emission perturbation responses within the source region for 6 selected regions (India, China, Europe, Germany, USA and Japan). The relative share of source regions to the pollutant levels within a given receptor region shows a lower inter-annual variability (typically between 2 and 6% for $PM_{2.5}$) than the absolute contributions.

4.3 Impact of the native TM5 grid resolution on pollutant concentration and SRs

FASST base concentrations and SRs have been derived at a 1°x1° resolution which is a relatively fine grid for a global model, but still not optimal for population exposure estimates and health impact assessments. Previous studies have documented the impact of grid resolution on pollutant concentrations. The effect of higher grid resolution in global models is in general to decrease ozone exposure in polluted regions and to reduce O_3 long-range transport, while $PM_{2.5}$ exposure – mainly to primary species - increases (Fenech et al., 2018; Li et al., 2016; Punger and West, 2013). Without attempting a detailed analysis, a comparison of TM5 available output for $PM_{2.5}$ and O_3 at 6°x4°, 3°x2° and 1°x1° resolution confirms these findings, as illustrated in Fig. S2.6 of the SI. Although FASST is expected to better represent population exposure to pollutants than coarser resolution models, a resolution of 1°x1° may not adequately capture urban scale pollutant levels and gradients when the urban area occupies only a fraction of the grid cell. The developed sub-grid parameterization for $PM_{2.5}$, providing an order-of-magnitude correction which is consistent with a high-resolution satellite product, is subject to improvement and to extension to other primary pollutants (NO_2, e.g. Kiesewetter et al., 2014, 2015) and O_3. To our knowledge a workable parametrization to quantify the impact of sub-grid O_3 processes on population exposure – in particular titration due to local high NO_x concentrations in urban areas - has not been addressed in global air quality models.

The impact of grid resolution on the within-region source-receptor coefficients can be significant, in particular for polluted regions where the coarse resolution includes ocean surface, like Japan. Table S2.3 in the SI shows as an example within-region and long-range SR coefficients for receptor regions Germany, USA and Japan. A higher grid resolution increases the within-region response and decreases the contribution of long–range transport (where the contribution of China to nearby Japan behaves as a within-region perturbation). In the case of Japan, the within-region $PM_{2.5}$ response magnitude increases with a factor of 3, and the sign of the within-region O_3 response is reversed when passing from 6°x4° to higher resolution. Also over the USA, the population-weighted within-region response sensitivity upon NO_x perturbation increases with a factor of 5. Further, we find that in titration regimes, the magnitude of the O_3 response to NO_x emissions increases with resolution (i.e. ozone increases more when NO_x is reduced using a fine resolution) whereas the in-region ozone response is reduced in non-titration regimes (India and China, Fig. 2.7d). These indicative results are in line with more detailed studies (e.g. Wild and Prather, 2006).

2) Also I found little mention of how the fixed meteorological year of 2001 could potentially impact the prediction of pollutants in the future i.e. how would climate change affect predictions of future pollutants?
Also the basis for the radiative forcing calculations is from a fixed meteorological year of 2001 and could have implications for the future calculation of effects. A more detailed mention of these issues would be good, perhaps in Section 4.

REPLY: This is an issue raised by both reviewers. We agree with the reviewer that the year 2001 meteorology is somewhat outdated. The perturbation runs for constructing the SR library of FASST were performed with the TM5 model set-up defined in the first phase of HTAP1 (during the period 2008 – 2011) and because of the computational costs, an update with more recent meteorology was not possible (TM5 is not taking part in HTAP2 where meteorological year 2010 has been used). A systematic check of the representativeness of this particular year for each of the FASST regions is beyond the scope of this study, in the first place because FASST is considered to be a screening tool focussing on impacts of emission changes. However we have substantially extended the discussion on the use of a single meteorological year.

CHANGES TO MANUSCRIPT: *Added to Section 2.1 P5 L10*

Meteorological fields are obtained from the ECMWF operational forecast representative for the year 2001. The implications of using a single meteorological year will be discussed in section 4.2.

Discussion section 4.2 added as included above.

3) TM5-FASST and the validation of it using TM5 simulations have all been conducted using emissions inventory for the year 2000 as a baseline along with 20% perturbations from this base. How appropriate is it to use a base year of 2000 for validation purposes given the large recent changes in emissions over the last 10-15 years, particularly over East Asia where some emissions have changed by >20%. What impact would using more up to date emissions in the base scenario have the calculated source-receptor coefficients and would it significantly affect the magnitude of future predictions? It would be useful to provide information on how recent changes in emissions could impact TM5-FASST.

REPLY: This is certainly an issue of concern, but at the same time difficult to address in a quantitative way. Although an independent set of SR simulations departing from a different reference scenario is not available, in the manuscript we included a validation of the linear scaling approach beyond the -20% perturbation, based on a number of additional perturbation simulations with TM5 for selected key regions, including East-Asia. In these test cases, the emissions of individual precursors where decreased by -80% relative to the reference emissions of the year 2000, while other precursors were kept at the year 2000 emissions. These simulations are not exactly testing the emission-response sensitivity for a different reference case, but they do provide a validation of the linear approach.
A second validation method, discussed as well in the paper, uses exactly emission scenarios that are strongly different from the reference year 2000 case for all precursors simultaneously (i.e. GEA FLE2030 and MIT2030 scenarios), where FASST uses the sensitivities based on year 2000 and compares the outcome with TM5, to some extend addressing the issue raised by the reviewer. The magnitudes of these emission changes are representative for more recent scenarios. A general observation is that FASST somewhat over-predicts resulting O3 and $PM_{2.5}$ concentrations (compared to the full TM5 model) for large (i.e. greater than say 50%) emission perturbations in either direction, but this does not compromise its usefulness as a tool to explore air pollution scenarios in a multi-pollutant/multi-impact framework. Indeed, as demonstrated in section 3.3, regional key features and trends, as well as inter-regional differences or similarities resulting from future RCP scenarios up to 2050 are reproduced within the variability of the ACCMIP air quality model ensemble.
We dedicate more discussion on these results, including a more systematic statistical analysis of the performance of FASST versus TM5. In the final discussion we refer to the new round of perturbation simulations performed in the frame of HTAP2.

CHANGES TO MANUSCRIPT: *we made substantial edits to the whole of section 3.2*

3.2: TM5-FASST_v0 versus TM5 for future emission scenarios

In this section we evaluate different combinations of precursor emission changes relative to the base scenario in a global framework. We take advantage of available TM5 simulations for a set of global emission scenarios which differ significantly in magnitude from the FASST base simulation, and as such provide a challenging test case to the application of the linear source-receptor relationships used in TM5-FASST. We assume that the full TM5 model provides valid evaluations of emission scenarios, and we test to what extent these simulations can be reproduced by the linear combinations of SRs implemented in the TM5-FASST_v0 model.
We use a set of selected policy scenarios prepared with the MESSAGE integrated assessment model in the frame of the Global Energy Assessment GEA (Rao et al., 2012, 2013; Riahi et al., 2012). These scenarios are the so called "frozen legislation" and "mitigation" emission variants for the year 2030 (named FLE2030, MIT2030 respectively), policy variants that describe two different policy assumptions on air pollution until 2030. These scenarios and there outcomes are described in detail in Rao et al. (2013), the scope of the present study is the inter-comparison between FASST and TM5 resulting pollutant concentration and exposure levels, as well as associated health impacts.
Major scenario features and emission characteristics are provided in section S8 of the SI. Table S8.1 shows the change in global emission strengths for the major precursors for both test scenarios, relative to the RCP2000 base, aggregated to the FASST 'master zoom' regions listed in Table S2.2. Emission changes for the selected scenarios mostly exceed the 20% emission perturbation amplitude from which the SRs were derived. Under the MIT2030 low emission scenario, all precursors and primary pollutants (except primary $PM_{2.5}$ in East-Asia and NH_3 in all regions) are showing a strong decrease compared to the RCP2000 reference scenario. The strongest decrease is seen in Europe (NO_x: -83%, SO_2: -93%, BC: -89%, primary $PM_{2.5}$ – 56%) while NH_3 is increasing by 14 to 46% across all regions. The FLE2030 scenario displays a global increase for all precursors, however with heterogeneous trends across regions. In Europe, North-America and Australia, the legislation in place, combined with use of less and cleaner fuels by 2030, leads to a decrease in pollutant emissions except for NH_3 and primary $PM_{2.5}$. On the other hand, very substantial emission increases are projected in East and South-East for BC, NO_x and primary $PM_{2.5}$. Anticipating possible linearity issues, we note that for both scenarios, in all regions, SO_2 and NO_x emissions are evolving in the same direction, although not always with similar relative changes, while NH_3 is always increasing, which may induce linearity issues in the ammonium-sulfate-nitrate system. Regarding O_3 metrics, NMVOC and NO_x are evolving in the same direction, but also here we observe possible issues due to a changing emission ratio (in particular in Russia and Asia).
We further note that not only the emission levels of these scenarios are different from the FASST base scenario (RCP year 2000), but also the spatial distribution of the emissions, at the resolution of grid cells, may differ from the reference set.
We use FASST to compute $PM_{2.5}$ and ozone concentrations applying Eq. (2), i.e. considering the FLE2030 and MIT2030 emission scenarios as a perturbation on the FASST reference emission set (RCP year 2000).
The scope of TM5-FASST is to evaluate on a regional basis the impacts of policies that affect emissions of short-lived air pollutants and their precursors. Hence we average the resulting O_3 and $PM_{2.5}$ concentration and O_3 exposure metric 6mDMA1 over the each of the 56 FASST regions and compare them with the averaged TM5 results for the same regions.
Further, in a policy impact analysis framework, the *change* in pollutant concentrations between two scenarios (e.g. between a reference and policy case) is often more relevant than the absolute concentrations. We therefore present absolute concentrations as well as the change (delta) between the two GEA scenarios, evaluating the benefit of a mitigation scenario versus the frozen legislation scenario.
Figure 8 shows the FASST versus TM5 regional scatter plots for absolute and delta population-weighted mean anthropogenic $PM_{2.5}$ for all 56 FASST receptor regions while the population-

weighted means over the 9 larger zoom areas are shown in Figure 9. Similarly annual mean population-weighted O_3 and 6mDMA1scatter plots are shown in Fig. 10, and the regional distribution in Fig. 11. The grid-cell statistics (mean, NMB, MB and R^2) over larger zoom areas are given in Tables 8 and 9 for $PM_{2.5}$ and 6mDMA1 respectively.

Figure 8 and Table 8 show that on a regional basis, the low emission scenario generally overestimates population-weighted $PM_{2.5}$ concentrations, with the highest negative bias in Europe and Asia, while the lowest deviation is found in Latin America and Africa. The agreement between FASST and TM5 is significantly better for the high emission scenario, in line with the findings in the previous section. As shown in Table 8, averaged over the larger zoom regions, we find that the relative deviation for $PM_{2.5}$ is within 11% for FLE2030, and within 28% for MIT2030, except for Europe where the (low) $PM_{2.5}$ concentration is overestimated by almost a factor of 2. The policy-relevant delta between the scenarios however is for all regions reproduced within 23%.

The ozone health metric 6mDMA1 is more scattered than annual mean ozone, and also here, as expected, the low emission scenario performs worse than the high emission one. Over larger zoom areas however the agreement is acceptable for both scenarios (FASST within 22% of TM5). Contrary to $PM_{2.5}$, the NMB for the delta 6mDMA1 between two scenarios is higher than the NMB on absolute concentrations, with a low bias for the delta metric of -38% and -45% for Europe and North-America respectively, and a high bias of 35 to 46% in Asia. However, the MB on the delta is of the same order or lower than the absolute concentrations (Table 9). This is a consequence of the fixed background ozone in the absolute concentration reducing the weight of the anthropogenic fraction in the relative error. Figures 9 and 11 provide a general picture of the performance of FASST: despite the obvious uncertainties and errors introduced with the FASST linear approximation, a consistent result emerges both for absolute concentrations from the individual scenarios as for the policy-relevant delta.

A major issue in air pollution or policy intervention impact assessments is the impact on human health; therefore we also evaluate the TM5-FASST outcome on air pollution premature mortalities with the TM5-based outcome, applying the same methodology on both TM5 and FASST outcomes. We evaluate mortalities from $PM_{2.5}$ using the IER functions (Burnett et al., 2014) and O_3 mortalities using the log-linear ER functions and RR's from Jerrett et al. (2009) respectively. Figure 12 ($PM_{2.5}$) and Fig. 13 (O_3) illustrate how FASST-computed mortalities compare to TM5, both as absolute numbers for each scenario, as well as the delta (i.e. the health benefit for MIT2030 relative to FLE2030). Regional differences in premature mortality numbers are mainly driven by population numbers. In line with the findings for the exposure metrics ($PM_{2.5}$ and 6mDMA1) FASST in general over-predicts the absolute mortality numbers, in particular in the low-emission case. For MIT2030, global $PM_{2.5}$ mortalities are overestimated by 19%, in Europe and North-America FASST even by 43%. In the FLE2030 case, we find a better agreement, with a global mortality over-prediction of 3% (for Europe and North-America 5% and 11% respectively). For the latter scenario, the highest deviation is found in Latin America (10 – 20%). O_3 mortalities are overestimated globally by 11% (7%) with regional agreement within 20% (14%) for MIT2030 (FLE2030). However, as shown by the error bars, the difference between FASST and TM5 is smaller than the uncertainty on the mortalities resulting from the uncertainty on RR's only. The potential health benefit of the mitigation versus the non-mitigation scenario (calculated as FLE2030 minus MIT2030 mortalities) is shown in Figs. 12c and 13c. Globally, FASST underestimates the reduction in global $PM_{2.5}$ mortalities by 17% with regional deviations ranging between -30% for Europe and North-America, and -12% for India. The global health benefit for ozone is underestimate by 2% for O_3, however as a net result of 11% overestimation in India and 12 to 59% underestimation in the other regions. The numbers corresponding to Figs. 12 and 13 are provided in Table S8.4 and S8.5 of the SI.

The error ranges presented here are obviously linked to the choice of the test scenarios and will for any particular scenario depend on the magnitude and the relative sign of the emission changes relative to RCP2000, but given the amplitude of the emission change for the currently two selected scenarios relative to RCP2000, these results support the usefulness of TM5-FASST as a tool for quick scenario screening.

4) In Section 2.1, P4, Line 12 the manuscript mentions about the advent of finer resolution global models nearing 1 ◦ x1 ◦ horizontal resolution. I think it would be good to make more comment on the applicability of the 1 ◦ x1 ◦ resolution when calculating country scale impacts. Is this resolution along with input information at similar resolution (e.g. emissions) sufficient to capture changes in pollutants at sub 100 km scales over countries such as the UK and Belgium/Luxembourg. I think that the urban adjustment of $PM_{2.5}$ is a suitable attempt at this but I think it would good to have some further comment on the issue of resolution and the limitations provided by other inputs at this resolution e.g. emissions and meteorology

REPLY: We agree with this critique, even if TM5 during the last decade or so has been amongst the global models with highest grid resolution that have made global studies on health impacts of air pollution. Further, the FASST source regions are defined such to include several gridboxes, e.g. Belgium/The Netherlands/Luxembourg are aggregated into a single region. We address the comment in the following ways:

1) Section 2.4, initially dedicated to the sub-grid adjustment for urban concentrations, has now been extended to include a quick analysis of the TM5 base simulation at resolution 6°x4°, 3°x2° and 1°x1° to illustrate the impact of resolution on concentrations and emission-concentration response sensitivities, with more detailed information and figures provided in the SI.
2) The paper already included a methodology to partly address the sub-grid gradients with a parametrized approach; in the connected annex S4 in the SI we now explicitly compare FASST $PM_{2.5}$ with a high-resolution satellite product.

CHANGES TO MANUSCRIPT:
Expanded section 2.1 P5 L26

With the introduction of massive parallel computing, however, this comparative advantage is now slowly disappearing, and global model resolutions of 1°x1° or finer are now becoming more common (see the model descriptions in this special issue, e.g. Liang et al., 2018). The model grid resolution influences the predicted pollutant concentrations as well as the estimated population exposure, especially near urban areas where strong gradients occur in population density and pollutant levels, which cannot be resolved by the 1°x1° resolution. In section 2.4 we describe a methodology to improve population $PM_{2.5}$ exposure estimates by applying sub-grid concentration adjustments based on high-resolution ancillary data. The bias introduced by model resolution affects as well computed SR matrices, e.g. off-setting the share of 'local' versus 'imported' pollution in a given receptor region. We will discuss this aspect more in detail in section 4.3.

(Section 4.3: see reply to comment 1)

5) Section 2.5 on health impacts provides a lot of details and is quite long compared to some others sections where most of the details are within a supplementary section. Also I found it a bit confusing to have two options for calculating $PM_{2.5}$ health effects: the log-linear and integrated exposure-response functions (IER). I assume the output from FASST is only provided from one (Figure 15)? The paragraph on page 10 Lines 8 to 13 does not seem to provide clarity on which method is preferred and could be re-worded. Therefore Section 2.5 could be potentially made more concise by removing the details on the log-linear method to the supplementary. This would allow the main text to focus more on the IER method by Burnett et al., (2014), which is the current methodology used within the Global Burden of Disease study.

REPLY: The reason for the relatively detailed description of health impacts calculations in TM5-FASST is that most users and publications tend to focus on this aspects- and because differences in methodologies are an important reason for differences in calculated health outcomes. We agree however that most of the description could move to the SI.
Most recently published global studies on health impacts of ambient air pollution use one of the two methodologies for $PM_{2.5}$ (i.e. log-lin and/or GBD) and both methodologies appear in WHO recommendations for Europe. We included both methods in the FASST output to facilitate comparison with other studies. The two calculations also provide an additional perspective on the uncertainty of the health impact outcome. (Upon request of Ref #2 we included an additional intercomparison of present-day mortalities with other studies, using both methodologies). However we agree with the reviewer that it was not clearly stated which method was used in Fig. 15 (now Fig. 17) – in this case, as the Silva study was based on GBD, we also used the result following the GBD methodology.

CHANGES TO MANUSCRIPT:
As suggested by the reviewer, we have moved a large part of the description of the health methodology (section 2.6) to the SI and kept only the GBD methodology in the main text, while mentioning that the tool includes the log-lin method as well.
We also mention now specifically in section 3.3.5 (Health impacts) that the methodology is based on GBD.

2.5 Health impacts

TM5-FASST provides output of annual mean $PM_{2.5}$ and O_3 health metrics (3-monthly and 6-monthly mean of daily maximum hourly O_3 (3mDMA1, 6mDMA1), and the sum of the maximal 8-hourly mean above a threshold of 35 ppbV (SOMO35) or without threshold (SOMO0), as well as annual mean NO_x and SO_2 concentrations at grid resolution of 1°x1°. These are the metrics consistent with underlying epidemiological studies (Jerrett et al., 2009; Krewski et al., 2009; Pope et al., 2002). The population-weighted pollutant exposure metrics grid maps, in combination with any consistent population grid map, are thus available for human health impact assessment. The TM5-FASST_v0 tool provides a set of standard methodologies, including default population and health statistics, to quantify the number of air quality-related premature deaths from $PM_{2.5}$ and O_3.
Health impacts from $PM_{2.5}$ are calculated as the number of annual premature mortalities from 5 causes of death, following the Global Burden of Disease methodology (Lim et al., 2012): ischemic heart disease (IHD), chronic obstructive pulmonary disease (COPD), stroke, lung cancer (LC) and acute lower respiratory airways infections (ALRI) whereas mortalities from exposure to O_3 are related to respiratory disease.
Cause-specific excess mortalities are calculated at grid cell level using a population-attributable fraction approach as described in Murray et al. (2003) from $\Delta Mort = m_0 \times AF \times Pop$, where m_0 is the baseline mortality rate for the exposed population, $AF = (RR-1)/RR$ is the fraction of total mortalities attributed to the risk factor (exposure to air pollution), RR = relative risk of death attributable to a change in population-weighted mean pollutant concentration, and Pop is the exposed population (adults ≥ 30 years old, except for ALRI for which infant population <5 years old was considered). RR for $PM_{2.5}$ exposure is calculated from the Integrated Exposure-Response functions (IER) developed by Burnett et al. (2014), and first applied in e.g. the Global Burden of Disease study (Lim et al., 2012).
In order to facilitate comparison with earlier studies, TM5-FASST provides as well mortality estimates based on a log-linear exposure response function $RR = \exp^{\beta\Delta PM_{2.5}}$ where β is the concentration–response factor (CRF; i.e., the estimated slope of the log-linear relation between concentration and mortality) and $\Delta PM_{2.5}$ is the change in concentration. More details on the health impact methodologies, as well as sources for currently implemented population and baseline mortality statistics and their projections in TM5-FASST_v0 are given in section S5 of the SI.

For O_3 exposure, $RR = e^{\beta(\Delta 6\text{mDMA1})}$, β is the concentration–response factor, and RR = 1.040 [95% confidence interval (CI): 1.013, 1.067] for a 10 ppb increase in 6mDMA1 according to Jerrett et al. (2009). We apply a default counterfactual concentration of 33.3 ppbV, the minimum 6mDMA1 exposure level in the Jerrett et al. (2009) epidemiological study.
We note that the coefficients in the IER functions used in the GBD assessments have been recently updated due to methodological improvements in the curve fitting, leading to generally higher RR and mortality estimates (Cohen et al.,2017; Forouzanfar et al., 2016). In particular, the theoretical minimum risk exposure level was assigned a uniform distribution of 2.4–5.9 μg/m^3 for $PM_{2.5}$, bounded by the minimum and fifth percentiles of exposure distributions from outdoor air pollution cohort studies, compared to the presently used range of 5.8 - 8.8 μg m^{-3} which would increase the health impact from $PM_{2.5}$ in relatively clean areas. Further, a recent health impact assessment (Malley et al., 2017), using updated RR estimate and exposure parameters from the epidemiological study by Turner et al. (2016), estimates 1.04–1.23 million respiratory deaths in adults attributable to O_3 exposure, compared with 0.40–0.55 million respiratory deaths attributable to O_3 exposure based on the earlier (Jerrett et al., 2009) risk estimate and parameters. These recent updates have not been included in the current version of TM5-FASST. Health impacts from exposure to other pollutants (NO_2, SO_2 for example) are currently not being evaluated in TM5-FASST-v0

In section 3.3.5 P27 L19

The analysis by Silva et al. (2016) used the same methodology implemented in FASST for estimating premature mortalities from $PM_{2.5}$ and O_3 (i.e. Burnett et al., 2014 as in the Global Burden of Disease study and Jerrett et al., 2009 respectively)

6) In section 3.1.1 when making a comparison of the additivity of emission perturbations for $PM_{2.5}$ individual changes for SO2, NOx and NH3 is shown on Figure 3 and 4 but in Figure 2 there is no effect from NH3 emissions. Whereas, in Figure S7.1 and S7.2 the 3 individual responses are shown along with the combined response on $PM_{2.5}$ (sum of all 3). However, the effect for combined emissions is only for SO2 and NOx in Figure 2 and 4 and does not include any addition from NH3. Why has the contribution from NH3 not been included within some of the combined emission changes in PM2.5? There seems to be a bit of inconsistency here, especially when considering that NH3 emissions can be important for NO3 aerosol formation.

REPLY: In first instance we have evaluated separately the 'additivity' and 'linearity' issues. Figure 2 demonstrates the additivity assumption of NOx and SO2 perturbations. This requires model simulations for (1) SO2 only perturbation (2) NOx only perturbation and (3) simultaneous NOx+SO2 perturbation, all of the same magnitude. For each of these perturbation experiments, the effect on SO4, NO3 and NH4 in $PM_{2.5}$ is available.
However due to lack of CPU resources, similar analyses for combined SO2+NH3 and NOx+NH3 perturbations have not been performed unfortunately. Because only separate NH3 perturbations are available we cannot provide the equivalent figures for these combinations. We therefore assume additivity for the combined perturbations of NH3 with SO2 and NOx respectively. To some extend one may argue that source regions of NH3 on the one hand, and SO2 and NOx on the other are less aligned, and that control strategies are different/independent hence simultaneous reductions are less pertinent, but we recognize this is a caveat in the FASST methodology.

CHANGES TO MANUSCRIPT:

3.1 Validation against the full TM5 model: additivity and linearity

We recall that the TM5-FASST computes concentrations and metrics based on a perturbation approach, i.e. the linearization applies only on the difference between scenario and reference emission. Therefore we focus on evaluating the perturbation response, i.e. the second term in the right hand side of Eq. 2.
The standard set of -20% emission perturbation simulations, available for all 56 continental source regions and constituting the kernel of TM5-FASST_v0 are simulations P1 (perturbation of SO_2, NO_x, BC and POM), P2 (SO_2 only), and P4 (NH_3 and NMVOC) shown in Table 2. Additional standard -20% perturbation experiments P3 (NO_x only) and P5 (NO_x and NMVOC), as well as an additional set of perturbation simulations P1' to P5' over the range [-80%, +100%], listed in Table S3 of the SI, have been performed for a limited selection of representative source regions (Europe, USA, China, India, Japan) due to limited CPU resources. For the same reason, no combined perturbation studies are available for (SO_2 + NH_3) and (NO_x + NH_3) for a systematic evaluation of additivity and linearity. The available [-80%, +100%] perturbations are used to validate the linearized reduced-form approach against the full TM5 model, exploring chemical feedback mechanisms (additivity) and extrapolation of the -20% response sensitivity towards larger emission perturbation magnitudes (linearity). This is in particular relevant for the NO_x - NMVOC - O_3 chemistry and for the secondary $PM_{2.5}$ components NO_3^- - SO_4^{2-} - NH_4^+. These mechanisms could also be important for organic aerosol, but we remind that in this study organic aerosol formation was parameterized as pseudo-emissions.

3.1.1Additivity and linearity of secondary inorganic $PM_{2.5}$ response:

Experiment P1, where BC, POM, SO_2 and NO_x emissions are simultaneously perturbed by -20% relative to base simulation P0, delivers SR matrices for primary components BC and POM, and a first-order approximation for the precursors SO_2 and NO_x whose emissions do not only affect SO_2 and NO_x gas concentrations but also lead to several secondary products (SO_2 forms ammonium sulfate, NO_x leads to O_3, ammonium nitrate). Experiment P2 perturbs SO_2 only, while experiment P3 perturbs NO_x only (in this latter case, to limit the computational cost, computed for a limited set of representative source regions only).
We first test the hypothesis that the $PM_{2.5}$ response to the combined (NO_x + SO_2) -20% perturbation (P1) can be approximated by the sum of the single precursor perturbations responses (P2 + P3). Figure 2 summarizes the resulting change in SO_4^{2-}, NO_3^-, NH_4^+ and total inorganic $PM_{2.5}$ respectively for the selected source regions. For Europe, the emission perturbations were applied over all European countries simultaneously, hence the responses are partly due to inter-regional transport from other countries. Following findings result from the perturbation experiments P1, P2 and P3:

1. Sulfate shows a minor response to NO_x emissions, and likewise nitrate responds only slightly to SO_2 emissions and both perturbations are additive. In general the response is one order of magnitude lower than the direct formation of SO_4^{2-} and NO_3^- from SO_2 and NO_x respectively (Fig. 2a, b);
2. NH_4 responds to NO_x and SO_2 emissions with comparable magnitudes and in an additive way (Fig. 2c);
3. The response of total sulfate, nitrate and ammonium to a combined NO_x and SO_2 -20% perturbation can be approximated by the sum of the responses to the individual perturbations, i.e. P1 ≈ P2+P3 (Fig. 2d). Scatterplots between P1 and P2+P3 for the regional averaged individual secondary products and total inorganic $PM_{2.5}$ are shown in Fig. S7.1 of the SI.

From the combined [SO_2+NO_x] perturbation (P1), and the separate SO_2 perturbation simulations (P2), both available for all source regions, the missing NO_x SR matrices have been gap-filled using (P1 – P2). By lack of simulations for combined (SO_2 + NH_3) or (NO_x + NH_3) perturbations we assume additivity for simultaneous NH_3, SO_2 and NO_x perturbations, i.e. the response is computed from a linear combination of P2, P3 and P4.

7) Within section 3 on the evaluation of TM5-FASST numerous references are made to the ability of FASST to predict TM5 concentrations or other metrics using the gradient of the straight line fit as an estimate of bias. I have noticed a couple of times in the text where FASST is stated to over or under estimate the comparison but the details in the figure do not agree with this statement, which could be due to the use of the gradient. I think that a more appropriate bias statistic such as normalised mean bias (or something similar) could be used to provide an evaluation of FASST rather than this simple linear fit. This occurs throughout Section 3 and please check that all comments are appropriate to the relevant figures.

REPLY:
This point is well taken, the slope of the fit was indeed not the most appropriate choice for evaluating the performance of FASST. We have omitted the linear fit in the figures, and leave only the 1:1 line as a reference. Instead we have calculated Normalized Mean Bias (NMB), Mean Bias (MB) and correlation coefficient as validation metrics in a consistent way across sections 3.1 and 3.2 when compare FASST to the full TM5 model, where

NMB = $\left(\overline{FASST - TM5}\right)/\overline{TM5}$

MB = $\left(\overline{FASST - TM5}\right)$

$\bar{Y}$ = average of all grid cells in region

Further, in section 3.1 (linearization error under strong emission perturbations) we focus on evaluating the perturbation term (delta), putting additional statistics on the total concentrations in the SI. In section 3.2 (comparison with high/low GEA emission scenarios) we show and discuss both totals for individual scenarios and deltas in the main text.

CHANGES TO MANUSCRIPT;
New discussion in section 3.1.2 P21 L6

Figure 7 illustrates the performance of the TM5-FASST approach versus TM5 for regional-mean annual mean ozone, health exposure metric 6mDMA1 (both evaluated as population-weighted mean), and for the crop-relevant exposure metrics AOT40 and M12 (both evaluated as area-weighted mean) over the extended emission perturbation range. In most cases the response (i.e. the *change* between base and perturbed case) to emission perturbations lies above the 1:1 line across the 4 metrics, indicating that FASST tends to over-predict the resulting metric (as a sum of base concentration and perturbation). Of the four presented metrics, AOT40 is clearly the least robust one, which can be expected for a threshold-based metric that has been linearized. Tables 5 to 7 give the statistical metrics for the grid-to-grid comparison of the perturbation term between FASST and TM5 for the health exposure metric 6mDMA1, and crop exposure metrics AOT40 and M12 respectively. Statistical metrics for the total absolute concentrations (base concentration + perturbation term) are given in Tables S7.2 to S7.4 in the SI. As anticipated, the NO_x-only perturbation terms are showing the highest deviation, in particular for a doubling of emissions, however combined NO_x-NMVOC perturbations are reproduced fairly well for all regions, staying within 33% for a -80% perturbation for all 3 exposure metrics, and within 38% for an emission doubling for 6mDMA1 and M12, while the AOT40 metric is overestimated by 76 to 126% for emission doubling. The total resulting concentration over the entire perturbation range for single and combined NO_x and NMVOC perturbation agrees within 5% for 6mDMA1 and M12, and within 64% for AOT40. The mean bias is positive for both perturbations, for all metrics and over all analysed regions, except for crop metric M12 under a doubling of NMVOC emissions over Europe showing a small negative bias. The deviations for individual European receptor regions under single and combined NMVOC and NO_x perturbations for health and crop exposure metrics are shown in Figs. S7.4 to S7.6 of the SI.

8) Within Section 3 a comparison has been made with air pollutant concentrations, health and climate metrics. However, no comparison has been made to other studies on the crop relevant metrics. The comparison of crop relevant metrics seems to have been excluded from the comparison. Is it possible to compare the results from FASST to other studies that have looked at the air pollution impact on crops to provide some evaluation of these metrics?

REPLY: We now include an intercomparison with a study on present-day global and regional crop losses

CHANGES TO MANUSCRIPT:
New section 3.3.6

3.3.6 Present day O3 – crop losses

Avnery et al. (2011) evaluate year 2000 global and regional O_3-induced crop losses for wheat, maize and soy bean, based on the same crop ozone exposure metrics as used in FASST, obtained with a global chemical transport model at 2.8°x2.8° resolution. Figure 18 compares their results (in terms of relative yield loss) with FASST (TM5) results based on RCP year 2000 for the globe and 3 selected key regions (Europe, North-America and East Asia). Despite the less-robust quantification of crop impacts from O_3 in a linearized reduced-form model set-up, we find that FASST reproduces the major features and trends across regions and crop varieties. Differences may be attributed to a variety of factors, including model resolution, model O3 chemistry processes, emissions, definition of crop growing season and crop spatial distribution.
And new Figure 18:

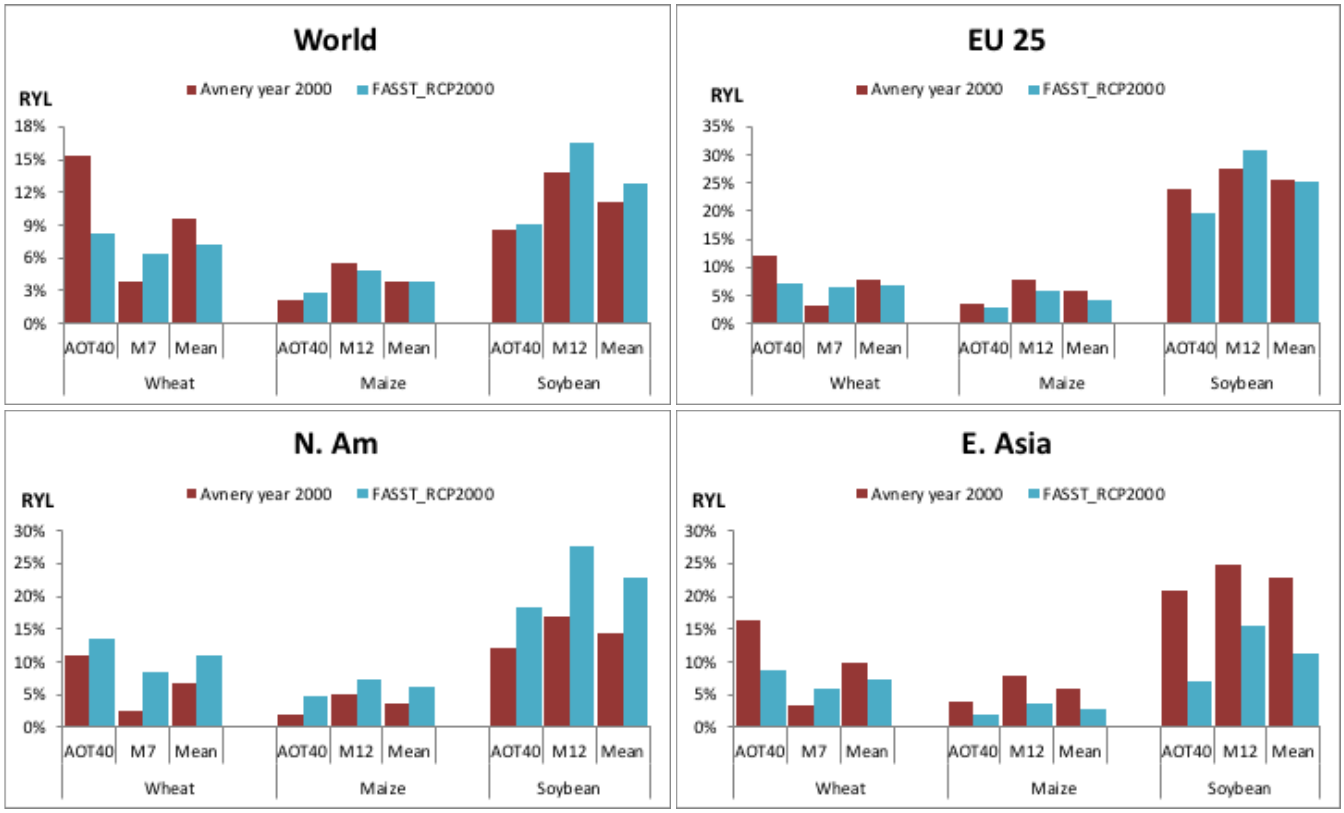

Figure 18:Year 2000 global and regional ozone-induced relative yield losses for 3 major crops, from Avnery et al. (2011) and from TM5-FASST (RCP year 2000), estimated from the 2 common exposure metrics M7 and AOT40 (see text), as well as the mean of both.

9) Please could the author make sure that all the equations provided within the manuscript are appropriately numbered. It appears that some have been but not all.

REPLY: OK, done

Minor Specific Comments:

10) Section 2.1, P3, Line 12 – Brackets needed round O3 as first time defined as ozone

REPLY: OK done

11) Section 2.1, P3, line 14 – When describing the particulate matter components I think some mention needs to be made here about Secondary Organic Aerosol (SOA). I think this comes later in the manuscript (section 2.3 P6) but I feel it would also be worth mentioning here with the initial model description

REPLY: Agree

CHANGES TO MANUSCRIPT: *P4 L16*

Biogenic secondary aerosol (BSOA) was included following the AEROCOM recommendation (Dentener et al., 2006; Kanakidou et al., 2005) which parameterized BSOA formation from natural VOC emissions as a fixed fraction of the primary emissions. The relative fraction compared to the anthropogenic POM emissions varied spatially, with a higher contribution in regions were the emissions of terpene emissions were higher.
SOA from anthropogenic emission was not explicitly included in the current simulations.

And to the discussion: P31 L17

The omission of secondary organic PM in TM5 is estimated to introduce a low bias in the base concentration of the order of 0.1 μg m^{-3} as global mean however with regional levels in Central Europe and China up to 1 μg m^{-3} in areas where levels of primary organic matter are reaching 20 μg m^{-3} (Farina et al., 2010) indicating a relatively low contribution of SOA to total $PM_{2.5}$.

12) Section 2.1, P3, Line 26 – 'Although for most health and ecosystem impacts only the surface level fields are required, base simulation and perturbed pollutants concentrations were calculated and stored for the 25 vertical levels of the model as monthly means, and some air quality-relevant parameters as hourly or daily fields.' – I think some mention of the fact that to calculate climate relevant impacts requires 3D information of constituents and not just surface fields

REPLY: Agree

CHANGES TO MANUSCRIPT: *changed the relevant phrase to: (P5 L7)*

Although for most health and ecosystem impacts only the surface level fields are required, climate metrics (e.g. radiative forcing) require the full vertical column and profile information. Therefore base simulation and perturbed pollutant concentrations were calculated and stored for the 25 vertical levels of the model as monthly means, and some air quality-relevant parameters as hourly or daily fields.

13) Section 2.3, P4, Line 25 – reference should be made to the underlying effects of the particular meteorological year used i.e. 2001 in this case.

REPLY: We do not fully agree that this addition would fit in here as the phrase describes a general feature of AQ-SRM. However we added it in the 3th par where TM5-FASST_v0 is introduced.

CHANGES TO MANUSCRIPT: *P6 L27*

In the current version v0 of TM5-FASST the emission-concentration relationship is locally approximated by a linear function expressing the change in pollutant concentration in the receptor region upon a change in precursor emissions in the source region with the generic form $dC_y = SRC \times dE_x$ where dC_y equals the change in the pollutant concentration compared to a reference concentration in receptor region *y*, dE_x is the change in precursor emission compared to a reference emission in source region *x*, and SRC the source-receptor coefficient for the specific compound and source-receptor pair – in this case emulating atmospheric processes linked to the meteorology in 2001.

14) Section 2.3, P5, Line 4 – '(Where j =i in the case of a primary component)' – maybe this could be changed to '(where the concentration of a primary pollutants is directly related to its emission)'

REPLY: We intend here that a primary component does not change chemically after its emission.

CHANGES TO MANUSCRIPT: *As the phrase is rather redundant we removed it.*

15) Section 2.3, P6, Lines 1 – 7 – There seems to be confusion between the labelling of emitted precursors and concentrations of components as in this section they both seemed to have been referred to as j. Please clarify which letter is meant to represent each

REPLY: This was indeed wrongly indexed, thanks for spotting.

CHANGES TO MANUSCRIPT*: P7 L29*

For each receptor point *y* (i.e. each model vertical level 1°x1° grid cell), the change in concentration of component *j* in receptor *y* resulting from a -20% perturbation of emitted precursor *i* in source region *x*, is expressed by a unique SR coefficient $A_{ij}[x,y]$:

$A_{ij}[x,y] = \frac{\Delta C_j(y)}{\Delta E_i(x)}$ with $\Delta E_i(x)$=$0.2E_{i,base}(x)$ (1)

The total concentration of component *j* in receptor region *y*, resulting from arbitrary emissions of *all* n_i precursors *i* at *all* n_x source regions *x,* is obtained as a perturbation on the base-simulation concentration, by summing up all the respective SR coefficients scaled with the actual emission perturbation:

$C_j(y) = C_{j,base}(y) + \sum_{k=1}^{n_x}\sum_{i=1}^{n_i} A_{ij}[x_k,y] \cdot \left[E_i(x_k) - E_{i,base}(x_k)\right]$ (2)

16) Section 2.3, P6, Equation 1 – Are these Source receptor coefficients calculated on the monthly or annual response between the precursor emission and pollutant? This needs to be stated within the description of the equation.

REPLY: Emission perturbations are implemented on annual basis, and the change in the source-receptor pollutant concentrations are evaluated on an annual basis as well. However some exposure metrics are based on seasonal values (e.g. crop growing season, human exposure to O3 during highest 6 monthly mean of hourly maximum values). We extended the paragraph, including as well additional information on the treatment of residual water in $PM_{2.5}$ to address an issue raised by Ref #2.

CHANGES TO MANUSCRIPT: *added phrase after Eq. 1(P8 L1)*

In the present version TM5-FASST_v0, the SR coefficients for pollutant concentrations are stored as annual mean responses to annual emission changes. Individual $PM_{2.5}$ components SRs are stored as dry mass ($\mu g\ m^{-3}$). $PM_{2.5}$ residual water at 35% is optionally calculated a posteriori for sensitivity studies, assuming mass growth factors for ammonium salts of 1.27 (Tang, 1996) and for sea-salt of 1.15 (Ming and Russell, 2001). The presence of residual water in $PM_{2.5}$ is not irrelevant: epidemiological studies establishing $PM_{2.5}$ exposure-response functions are commonly based on monitoring data of gravimetrically determined PM2.5, for which measurement protocols foresee filter conditioning at 30 – 50% RH. Therefore, although most health impact modelling studies consider dry $PM_{2.5}$ mass, the residual water fraction should in principle be included in modelled PM2.5.
We also established SR matrices linking annual emissions to specific O_3 exposure metrics that are based on seasonal or hourly O_3 concentrations (e.g. crop exposure metrics based on daytime ozone during crop growing season, human exposure to O_3 during highest 6 monthly mean of hourly maximum values).

And deleted the phrase below Eq (3):

"In TM5-FASST_v0 the monthly perturbations are aggregated to annual emission-concentration SR matrices, as the health, climate and vegetation impact metrics used in this version are also aggregated to annual values."

17) Section 2.3, P6, Line 21 - 24 – It is unclear to me how secondary organic aerosol (SOA) is included within the TM5-FASST tool as a component of PM2.5. Does it form part of the POM and what fraction of the primary emissions are used

REPLY: This is partly addressed in our reply to comment 10). Further we specify now that the perturbation simulations are made for anthropogenic components only.

CHANGES TO MANUSCRIPT: *Paragraph was modified as follows (P7 L14)*

The SR matrices, describing the concentration response in each receptor upon a change in emissions in each source region, have been derived from a set of simulations with the full chemical transport model TM5 by applying -20% emission perturbations for each of the 56 defined source regions (plus shipping and aviation), for all relevant anthropogenic precursor components, in comparison to a set of unperturbed simulations, hereafter denoted as 'base simulations'. Emissions from biogenic organic components were included as a spatial/temporally varying component, but did not vary in the model emission sensitivity simulations. Consequently absolute concentrations of BSOA were identical across base and perturbation simulations and no SR coefficients are available.

18) Section 2.3, P7, Line 3 – The combination of emissions perturbation scenarios is given in Table 2. Did the base simulation not conduct emission perturbation scenarios for all 56 continental regions? I thought that this would have been essential to enable to the calculation of changes in concentrations in TM5 but Table 2 does not seem to imply this. Clarification required

REPLY: We agree that the phrase is formulated confusingly and deserves more clarification. The purpose of the perturbation simulations is indeed to obtain SR matrices for each precursor, and for each of the source regions, but it was not required to run all individual perturbations for all

regions. Table 2 explains in brief the purpose of each simulation and section 3.1 explains in detail how the various simulations are combined to get to the full set.

CHANGES TO MANUSCRIPT: *Added to section 2.3: (P9 L8)*

The -20% perturbation simulations were performed for the combination of precursors given in Table 2, with P0 the unperturbed reference simulation, and P1 through P5 -20% perturbations for combined or single precursors. Due to limited CPU availability, precursors that are expected not to interact chemically are perturbed simultaneously, with P1 combining SO_2, NO_x, BC, and POM and P4 combining NH_3 and NMVOC. P1 and P4 were computed for each of the 56 continental source regions plus shipping (P1 and P4) and aviation (P1). Additionally, a SO_2-only perturbation was computed for all individual source regions and shipping (P2) and NO_x-only for a selection of key source regions (P3). Finally a set of combined NO_x + NMVOC perturbation simulations (P5) was performed for a set of key regions.
For a limited set of representative source regions, an additional wider range of emission perturbations P_i' [-80% to +100%] has been applied to evaluate possible non-linearities in the emission-concentration relationships. The list of these additional perturbation simulations is given in Table S3 of the SI. In section 3.1 we explain how this set of perturbation runs is combined into FASST to obtain a complete set of source-receptor matrices for each precursor and source region.

Modified Section 3.1

3.1 Validation against the full TM5 model: additivity and linearity
We recall that the TM5-FASST computes concentrations and metrics based on a perturbation approach, i.e. the linearization applies only on the difference between scenario and reference emission. Therefore we focus on evaluating the perturbation response, i.e. the second term in the right hand side of Eq. 2.
The standard set of -20% emission perturbation simulations, available for all 56 continental source regions and constituting the kernel of TM5-FASST_v0 are simulations P1 (perturbation of SO_2, NO_x, BC and POM), P2 (SO_2 only), and P4 (NH_3 and NMVOC) shown in Table 2. Additional standard -20% perturbation experiments P3 (NO_x only) and P5 (NO_x and NMVOC), as well as an additional set of perturbation simulations P1' to P5' over the range [-80%, +100%], listed in Table S3 of the SI, have been performed for a limited selection of representative source regions (Europe, USA, China, India, Japan) due to limited CPU resources. For the same reason, no combined perturbation studies are available for (SO_2 + NH_3) and (NO_x + NH_3) for a systematic evaluation of additivity and linearity. The available [-80%, +100%] perturbations are used to validate the linearized reduced-form approach against the full TM5 model, exploring chemical feedback mechanisms (additivity) and extrapolation of the -20% response sensitivity towards larger emission perturbation magnitudes (linearity). This is in particular relevant for the NO_x - NMVOC - O_3 chemistry and for the secondary $PM_{2.5}$ components NO_3^- - SO_4^{2-} - NH_4^+. These mechanisms could also be important for organic aerosol, but we remind that in this study organic aerosol formation was parameterized as pseudo-emissions.

19) Section 2.3, P7, Line 15 to 18 – The change in CH4 burden in TM5 from the HTAP1 perturbation simulations is stated as being an emission perturbation of 77 Tg/year. Could the authors provide information on how this was obtained

REPLY: The value comes from the assumption that the imposed CH_4 steady state concentration is the result of a balanced emission on the one hand and the chemical loss by oxidation by OH on the other hand (neglecting the lower-order losses to soil and stratosphere). As the TM5 model keeps track of the total amount of CH_4 oxidized, the implied change in emission is simply obtained from the difference in total amount of CH_4 oxidized in 1 year between the two runs.

We agree this could be explained better. In order to address a similar comment from Ref. #2 we have moved the details of the methodology to the SI, and modified the text as follows:

CHANGES TO MANUSCRIPT: *Main text, P9 L25:*

Annex S3 in the SI provides more details on the methodology applied to convert the CH_4 concentration perturbation into a CH_4 emission-based perturbation

Annex S3:

S3.1 CH_4 – O_3 source-receptor relations from HTAP1 perturbation experiments:
CH_4 emissions lead to a change in CH_4 concentrations with a perturbation response time of about 12 years. In order to avoid expensive transient computations, HTAP1 simulations SR1 and SR2 with prescribed fixed CH_4 concentrations (1760 ppb and 1408 ppb, see Dentener et al., 2010) were used to establish CH_4 – O_3 response sensitivities. Previous transient modeling studies have shown that a change in steady-state CH_4 abundance can be traced back to a sustained change in emissions, but the relation is not linear because an increase in CH_4 emissions removes an additional fraction of atmospheric OH (the major sink for CH_4) and prolongs the lifetime of CH_4 (Fiore et al., 2002, 2008; Prather et al., 2001).
In a steady-state situation, the CH_4 concentration is the result of balanced sources and sinks. In the HTAP1 experiments, keeping all other emissions constant, the change in the amount of CH_4 loss (mainly by OH oxidation with a lifetime of ca. 9 years, neglecting loss to soils and stratosphere with lifetimes of ca.160 and 120 years respectively (Prather et al., 2001)) under the prescribed change in CH_4 abundance should therefore be balanced by an equal and opposite source which we consider as an "effective emission". The amount of CH_4 oxidized by OH in one year being diagnosed by the model, the resulting difference between the reference and perturbation experiment of -77 Tg sets the balancing "effective" emission rate to 77Tg/yr, which is then used to normalize the resulting O_3 and O_3 metrics response to a CH_4 emission change.

The same perturbation experiments also allow us to establish the CH_4 self-feedback factor F describing the relation between a change in emission and the change in resulting steady-state concentration:

$$\frac{C_2}{C_1} = \left(\frac{E_2}{E_1}\right)^F \qquad \text{(S3.1)}$$

With CH_4 concentrations prescribed, CH_4 emissions were not included in the SR1 and SR2 experiments. The feedback factor F is derived from model-diagnosed respective CH_4 burdens (B) and total lifetimes (LT) as follows (Fiore et al., 2009; Wild and Prather, 2000):

$F=1/(1-s)$

$s = \partial\ln(LT) / \partial\ln(B)$

TM5 returns s = 0.33 which can be compared to a range of values between 0.25-and 0.31 in IPCC-TAR (Prather et al., 2001, Table 4.2) , resulting in a TM5-inherent calculated feedback factor F=1.5. This factor can be used to estimate the corresponding SR2-SR1 change in CH_4 emission in a second way. From Eq. S3.1 we find that a 20% decrease in CH_4 abundance corresponds to a 14% decrease in total CH_4 emissions. Kirschke et al. (2013) estimate total CH_4 emissions in the 2000s in the range 550 – 680 Tg yr^{-1} from which we obtain an estimated emission change between the HTAP SR1 and SR2 experiments in the range 77 – 95 Tg yr^{-1}, in line with our steady-state loss-balancing approach.

20) Section 2.3, P7, Lines 22 to 28 – It is stated that FASST does not include impacts on O3 from perturbations in CO emissions. I am not sure why this has not been included in the development of FASST along with other O3 precursor emissions of NOx and NMVOCs. Within this section it states that there is a dedicated CO emission perturbation experiment conducted with TM5 as part of HTAP1 available and that the impacts on O3 are not insignificant. Therefore I wonder why the information from the TM5 CO experiments have not been included previously within FASST?

REPLY: This is indeed a missing link in the TM5-FASST model which we hope to address in a future version of the tool. Also here, missing CPU resources did not allow for dedicated CO perturbation simulations in each of the 56 source regions. Indeed from HTAP1, source receptor relations between large rectangular source areas (not aligned with political borders and coast lines, and including ocean) are available but we did not attempt to remap those on the FASST 56 continental regions, given expected differences in CO lifetimes for emissions from these regions. With HTAP2 source regions better aligned with the FASST ones, there may be possibilities to rely on those in future developments. This caveat has been mentioned in the discussion.

21) Section 2.4, P8 – Maybe this section should be labelled as something like 'Urban Adjustments in $PM_{2.5}$ for Health Calculation' to better identify what is being done here. I am assuming that the adjusted $PM_{2.5}$ concentrations are only used within the calculation of health impacts

REPLY: Thank you for the suggestion – indeed this is relevant for the exposure of population. As this section now also includes a discussion on the impact of grid resolution (see reply to comment 4) we have modified the title

CHANGES TO MANUSCRIPT: *Title changed to:*

2.4 PM2.5 adjustments in urban regions for health impact evaluation

22) Section 2.4, P8, Lines 25 -26 – Is the CIESIN population dataset the default one used within FASST as this seems to have been used to calculate the default urban increment factors in Table S4.2? Might be worth included which one is recommended for use.

REPLY: The CIESIN dataset is the one with the highest resolution and therefore most suitable for a sub-grid correction. The 'default' regional increment factors are indeed based on CIESIN year 2000 data, but they are static and therefore do not change with scenario years. The public web tool always uses these default factors, but the (not-public) 'research version' has the option to include more appropriate population data sets.

CHANGES TO MANUSCRIPT: *we added the following phrase in the conclusion section: (P33 L30)*

This version offers the possibility to explore built-in as well as user-defined scenarios, using static default urban increment correction factors and crop production data. A more sophisticated in-house research version with gridded output and flexibility in the choice of gridded ancillary data (population grid maps, scenario-specific urban increment factors, crop distribution) is under continuous development and has been applied for the assessments listed in table S1.

23) Section 2.5, P9, Line 10 – Check definition of AF here as this does not match up with what is provided further down the page, just above line 20

REPLY: In fact it does correspond: 1-1/RR = (RR-1)/RR. But as the right-hand form is probably more legible we changed it to the latter. The part of the text above line 20 containing the larger equation has been moved to the SI following comment 4.

CHANGES TO MANUSCRIPT: *changed the phrase to: (P11 L21)*

… where m_0 is the baseline mortality rate for the exposed population, *AF = (RR-1)/RR* is the fraction of total mortalities attributed to the risk factor (exposure to air pollution)

24) Section 2.5, P10, Lines 17 to 24 – I think a comment is required here to state how the recent updates in the epidemiological evidence for health effects could impact on the predictions in FASST i.e. will they be cause an underestimate or overestimate.

REPLY: We have added a line to clarify the impact of the new parameter on the estimated health impact for PM2.5.

CHANGES TO MANUSCRIPT: *extended the phrase as follows: (P12 L6)*

In particular, the theoretical minimum risk exposure level was assigned a uniform distribution of 2.4–5.9 μg/m^3 for $PM_{2.5}$, bounded by the minimum and fifth percentiles of exposure distributions from outdoor air pollution cohort studies, compared to the presently used range of 5.8 - 8.8 μg/m^3 which would increase the health impact from $PM_{2.5}$ in relatively clean areas.

25) Section 2.6, P11, Line 14 – 'Both Mi metrics ...' should be changed to 'Both metrics (Mi) ...

REPLY: OK done

26) Section 2.6, P11, Line 15 – How is the growing season defined when calculating the crop metrics?

REPLY: As reported in the text, the growing seasons for the respective crops are retrieved from the gridded GAEZ data set. To clarify this more, we have extended the description of methodology related to the definition of the crop season.

CHANGES TO MANUSCRIPT: *modified section 2.6 as follows:*

2.6 Crop impacts

The methodology applied in TM5-FASST to calculate the impacts on four crop types (wheat, maize, rice, and soy bean) is based on Van Dingenen et al. (2009). In brief, TM5 base and -20% perturbation simulations of gridded crop O_3 exposure metrics (averaged or accumulated over the crop growing season) are overlaid with crop suitability grid maps to evaluate receptor region-averaged exposure metrics SR coefficients. Gridded crop data (length and centre of growing period, as well as a gridded crop-specific suitability index, based on average climate 1961 – 1990) have been updated compared to Van Dingenen et al. (2009), using the more recent and detailed Global Agro-Ecological Zones (GAEZ) data set (IIASA and FAO, 2012, available at http://www.gaez.iiasa.ac.at/).
Available crop ozone exposure metrics are 3-monthly accumulated ozone above 40 ppbV (AOT40) and seasonal mean 7 hr or 12 hr day-time ozone concentration (M7, M12) for which exposure-response functions are available from the literature (Mills et al., 2007; Wang and Mauzerall, 2004). Both metrics (M_i) are calculated as the 3-monthly mean daytime (09:00 – 15:59 for M7, 08:00 – 19:59 for M12) ozone concentration, evaluated over the 3 months centred on the midpoint of the location-dependent crop-growing season provided by the GAEZ data set.

Note that in the GAEZ methodology, the theoretical growing season is determined based on prevailing temperatures and water balance calculations for a reference crop, and can range between 0 and 365 days, however our approach always considers 3 months as the standard metric accumulation or averaging period.

27) Section 2.6, P11, Line 16 – RYL is defined as the crop relative yield. Should this be the relative yield loss? Also the coefficients a,b,c within the equation for RYL need more explanation

REPLY: indeed, "RYL" was wrongly positioned in the phrase. We have included a table with the values of the coefficients in the equations. While in the Weibull function the a and b parameters are pure mathematical shape coefficients, the c coefficients sets the lower threshold value for zero impact. We included this as well.

CHANGES TO MANUSCRIPT:

Modified following section (P12 L25):
Both metrics (M_i) are calculated as the 3-monthly mean daytime (09:00 – 15:59 for M7, 08:00 – 19:59 for M12) ozone concentration, evaluated over the 3 months centred on the midpoint of the location-dependent crop-growing season.
The crop relative yield loss (RYL) is calculated as linear function from AOT40 and from a Weibull-type exposure-response as a function of Mi:

$$RYL[AOT40] = a \times AOT40 \qquad (5)$$

$$\left.\begin{array}{ll} RYL(M_i) = 1 - \dfrac{exp\left[-\left(\frac{Mi}{a}\right)^b\right]}{exp\left[-\left(\frac{c}{a}\right)^b\right]} & M_i \geq c \\ RYL(M_i) = 0 & M_i < c \end{array}\right\} \qquad (6)$$

The parameter values in the exposure response functions are given in Table 3. Note that for $M_i = c$, $RYL = 0$ hence c is the lower M_i threshold for visible crop damage. Also here, the non-linear shape of the RYL(Mi) function requires the ΔRYL for 2 scenarios (S1, S2) being evaluated as RYL($M_{i,S2}$) – RYL ($M_{i,1}$), and not as RYL ($M_{i,S2}$- $M_{i,S1}$).

28) Section 2.7.1, P12, Lines 10 to 12 – Are these two sentences on the basic radiative properties of aerosols relevant? Including some text on the following lines would be good to discuss how the treatment of externally mixed aerosols alters the radiative forcing calculations when compared to internally mixed ones (Lesins et al., 2002; Klingmüller et al., 2014).

REPLY: We agree on the redundancy of the two sentences and removed them. We included a brief discussion on the impact of the introduced simplifications regarding mixing state as well as the use of integrated column burden instead of resolved vertical profiles. With respect to the mixing state we rather refer to Bond et al. (2013) who considered various additional processes affecting the BC absorption coefficient

CHANGES TO MANUSCRIPT: *Added following text: (P14 L15)*

Neglecting the aerosol mixing state and using column-integrated mass rather than vertical profiles introduces additional uncertainties in the resulting forcing efficiencies. Accounting for internal mixing may increase the BC absorption by 50 to 200% (Bond et al., 2013), while including the vertical profile would weaken BC forcing and increase SO4 forcing (Stjern et al.

2016). Further, the BC forcing contribution through the impact on snow and ice is not included, nor are semi- and indirect effects of BC on clouds. Our evaluation of pre-industrial to present radiative forcing in the validation section demonstrates that, in the context of the reduced-form FASST approach, the applied method however provides useful results..

29) Section 2.7.2, P12 – I think this sections needs to be made clearer. I am struggling to make the link between the output from FASST and the calculation of indirect aerosol forcing. How is done? What fields from FASST are used to calculate it? Needs to explain the methodology better for the reader.

REPLY: Apologies if the manuscript lacked clarity on this issue. Equation (7) explains how FASST SR matrices for radiative forcing are obtained: the change in forcings (both direct and indirect) for the perturbation experiments are computed from TM5-output using normalized forcing efficiencies. FASST then simply contains a SR coefficient to be multiplied with the emission change to obtain a forcing change. Sections 2.7.1 and 2.7.2. describe the underlying methodology in TM5. We have added some more clarification as follows:

CHANGES TO MANUSCRIPT: *modified section 2.7.3 as follows:*

2.7.3 Radiative forcing by O_3 and CH_4

Using TM5 output, indirect forcing is evaluated considering only the so far best studied first indirect effect, and using the method described by Boucher and Lohmann (1995). Fast feedbacks on cloud lifetimes and precipitation were not included in this off-line approach. This simplified method uses TM5 3D time-varying fields of SO4 concentrations, cloud liquid water content, and cloud cover (the latter from the parent ECMWF meteorological data). The parameterization uses the cloud information (liquid water content and cloud cover) from the driving ECMWF re-analysis data (year 2001). Fast feedbacks on cloud lifetimes and precipitation were not included in this off-line approach. The cloud droplet number concentrations and cloud droplet effective radius were calculated following Boucher and Lohmann (1995) separating continental and maritime clouds. The equations are given in section S6 of the SI. The global indirect forcing field associated with sulfate aerosols is shown in Fig. S6.1(d) of the SI. Indirect forcing by clouds remains however highly uncertain, and although FASST evaluates its magnitude, it is often not included in our analyses.

30) Section 2.7.2, P12, Line 29 – Add year used to meteorological data

REPLY: done (see previous comment)

31) Section 2.7.2, P12, Line 30 – missing word 'using' between after 'calculated'. Also it is probably worth stating here or in the supplementary section S6 the equations used to calculate cloud droplet number concentrations and cloud effective radius

REPLY: done

CHANGES TO MANUSCRIPT:
Following section was added to section S6 of the SI:

Indirect forcing:
The cloud droplet number concentrations (*CDNC*) were calculated using the following set of equations from Boucher and Lohmann (1995), separating continental and maritime clouds:

$$CDNC_{cont}^{St} = 10^{2.24+0.257log(m_{SO_4})}$$

$$CDNC_{cont}^{Cu} = 10^{2.54+0.186log(m_{SO_4})}$$
$$CDNC_{ocean} = 10^{2.06+0.48log(m_{SO_4})}$$

Following Boucher and Lohmann (1995), the cloud droplet effective radius is calculated from the mean volume cloud droplet radius:

$$r_e = 1.1\left(\frac{l\rho_{air}}{(4/3)\pi\rho_{water}CDNC}\right)^{1/3}$$

Where l = cloud liquid water content, ρ_{air} = air density, ρ_{water} = water density

32) Section 2.7.3, P13 – Like section 2.7.2. I think this section needs to be made clearer to highlight what output is being used from FASST to compute O_3 and CH_4 radiative forcings. There is a lot of details of what is included but I struggled to follow the basic principle of FASST output + forcing efficiency = radiative forcing. I think the description of what is done in FASST should come first at the start of this paragraph and then follow with the description of what it takes account of.

REPLY: We apologize for the lack of clarity. The section was indeed not very clear in explaining the methodology used in TM5 and how this is transferred into FASST. We have modified the introductory part of section 2.7 to explain the general approach: TM5 provides radiative forcing output from a built-in methodology, and the forcing SRs in FASST are simply based on emission-normalized delta's between base and perturbation experiments. The subsequent sections then explain in more detail how forcing is calculated in TM5.
Further we have shortened section 2.7.3 and moved the details of the methodology to the SI (new section S6.2)

CHANGES TO MANUSCRIPT: *we modified the introductory part of the section and the section addressing radiative forcing by O3 and CH4 as follows:*

2.7 Climate metrics

We make use of the available 3D aerosol and O_3 fields in the -20% emission perturbation simulations with TM5 to derive the change in global forcing for each of the perturbed emitted precursors. The region-to-global radiative forcing SR for precursor *j*, emitted from region *k*, is calculated as the emission-normalized change in global radiative forcing between the TM5 base and the corresponding -20% emission perturbation experiment:

$$SR\_RF_k^j = \frac{RF\_PERT[j,k] - RF\_BASE}{0.2E_k^j}\ [\mathrm{W/m^2}]/[\mathrm{kg/yr}] \qquad (7)$$

where RF_PERT and RF_BASE are the TM5 global radiative forcings for the perturbation and base simulations respectively, and E_k^j is the annual base emission of precursor *j* from region *k*. For each emitted pollutant (primary and secondary) the resulting normalized global forcing responses are then further used to calculate the global warming potential (GWP) and global temperature potential (GTP) for a series of time horizons H. In this way, a set of climate metrics is calculated with a consistent methodology as the air quality metrics, health and ecosystem impacts calculated from the concentration and deposition fields. In this section we describe in more detail the applied methodologies in TM5 to obtain the radiative forcing from aerosols, clouds and gases, as well as the derivation of the GWP and GTP metrics.

(…)

2.7.3 Radiative forcing by O_3 and CH_4

Using TM5 output, radiative forcing (RF) by ozone is approximated using the forcing efficiencies obtained by the STOCHEM model as described in Dentener et al. (2005), normalized by the ozone columns obtained in that study. Here we use annual averaged forcing based on the

RF computations provided as monthly averages by D. Stevenson (personal communication, 2004). The radiative transfer model was based on Edwards and Slingo (1996). These forcings account for stratospheric adjustment, assuming the fixed dynamical heating approximation, which reduces instantaneous forcings by ~22%.
For CH_4 the RF associated with the base simulation was taken from the equations in the IPCC-Third Assessment Report (TAR) (Table 6.2 of Ramaswamy et al., 2001). Using the HTAP1 calculated relationship between CH4 concentration and emission, and the same equations, we evaluated a globally uniform value of 2.5 mW/m^2 per Tg CH_4 emitted. (Dentener et al., 2010). It includes both the direct CH_4 greenhouse gas (GHG) forcing (1.8 mW/m^2) as well as the long-term feedback of CH_4 on hemispheric O_3 (0.7 mW/m^2).
From the TM5 perturbation experiments we derive as well region-to-global radiative forcing SRs for precursors (NO_x, NMVOC, CO and SO_2) through their feedback on the CH_4 lifetime and subsequently on long-term hemispheric O_3 levels. Hence, the greenhouse gas radiative forcing contribution of each ozone precursor consists of 3 components: a direct effect through the production of O_3, a contribution by a change in CH_4 through modified OH levels (including a self-feedback factor accounting for the modified CH_4 lifetime), and a long-term contribution via the feedback of CH_4 on hemispheric ozone. The details of the applied methodology are given in section S6.2 of the SI.
In its current version, TM5-FASST_v0 provides the steady-state concentrations and forcing response of the long-term O_3 and CH_4 feedback of sustained precursor emissions, i.e. it does not include transient computations that take into account the time lag between emission and establishment of the steady-state concentration of the long-term O_3 and CH_4 responses.

And in the SI:

S6.2 Secondary forcing feedbacks of O3 precursors on CH_4 and background O_3
Emissions of short-lived species (NOx, NMVOC, CO, SO_2) influence the atmospheric OH burden and therefore the CH_4 atmospheric lifetime, which in turn contributes to long-term change in CH4 and background ozone. Hence, the total forcing contribution from O3 precursors consists of a short-term direct contribution from immediate O3 formation (S-O3), and secondary contributions from CH_4 (I-CH4) and a long-term feedback from this CH_4 on background O3 (M-O3).
We apply the formulation by (Fiore et al., 2009; Prather et al., 2001; West et al., 2007) to calculate the secondary change in steady-state CH_4 from SLS emissions, using the TM5 perturbation experiments for FASST (see section S3). TM5 diagnoses the CH_4 loss by oxidation for reference and perturbation run (where the emissions of SLS are decreased with -20%), from which we calculate the CH_4 oxidation lifetime ratio between reference and perturbation:

$$\frac{LT_P}{LT_{Ref}} = \frac{CH4\_ox_P}{CH4\_ox_{Ref}} \qquad [S6.5]$$

Where LT is the CH4 lifetime against loss by OH oxidation, and $CH4\_ox$ = the amount (Tg) of CH4 oxidized.
The new steady-state methane concentration M due to the changing lifetime from perturbation experiment P, induced by O_3 precursor emissions follows from (Fiore et al., 2008, 2009; Wild and Prather, 2000):
$M = M_0 \times \left(\frac{LT_P}{LT_{ref}}\right)^F$ where $M_0 = 1760$ ppb, the reference CH_4 concentration and F = 1.5, determined from the HTAP1 CH4 perturbation experiments, as described in section S3.

The change in CH_4 forcing (I-CH4) associated with the change to the new steady-state concentration is obtained from IPCC AR5 equations:

$$\Delta F = \alpha\left(\sqrt{M} - \sqrt{M_0}\right) - \left(f(M, N_0) - f(M_0, N_0)\right) \quad [S6.6]$$

$$f(M, N) = 0.47ln[1 + 2.01 \times 10^{-5}(MN)^{0.75} + 5.31 \times 10^{-15}M(MN)^{1.52}] \quad [S6.7]$$

Where M, M_0 = CH_4 concentration in ppb, N_0 = N_2O (=320 ppb)

The associated long-term O3 forcing (M-O3) per Tg precursor emitted is obtained by scaling linearly the change in O_3 forcing obtained in the HTAP1 CH_4 perturbation simulation (SR2–SR1), with the change in CH_4 obtained above, and normalizing by the precursor emission change (Fiore et al., 2009)

$$\Delta F = \frac{\Delta F_{O3}[SR2-SR1]}{M_{SR2}-M_{SR1}}(M - M_0) \quad [S6.8]$$

The response of CH_4 and O_3 forcing to CO emission changes (for which no regional TM5-FASST perturbation model simulations were performed) was taken from TM5-CTM simulations performed for the HTAP1 assessment (Dentener et al., 2010) using the average forcing efficiency for North America, Europe, South-Asia and East-Asia. For regions not covered by the HTAP1 regions, the HTAP1 rest-of-the-world forcing efficiency was used.
The resulting region-to-globe emission-based forcing efficiencies are given in Tables S6.2 to S6.5 for aerosols, CO, CH_4 and other O_3 precursors respectively.

33) Section 2.7.3, P13, Lines 4 to 6 – How do these STOCHEM calculations compare to the ACCMIP multi-model mean and is it still appropriate?

REPLY: We have not made ourselves the comparison between STOCHEM and ACCMIP normalized O3 radiative forcings. However Stevenson et al. (2013)calculated a global normalized RF of 42 mWm^{-2} DU^{-1}, while two other model studies find values of about 36 mWm^{-2} DU^{-1}. In this study a value 30 mWm^{-2} DU^{-1} was found, broadly in line with the global numbers above. The results of Stevenson et al. (2013) were not available when the RF module was developed, and indeed updating the radiative transfer code, including ozone vertical profiles (instead of using fixed ozone columns) would be obvious candidates for improvement.

34) Section 2.7.3, P13, Line 32 – For regions not covered by the major HTAP1 source could the 'rest of the world' CO forcing efficiency not be used from Table S6.3 rather than a global average?

REPLY: This is indeed a correct observation; we have corrected the text and the values in Table S6.3.

35) Section 2.7.4, P14, Line 7 – Are the emission based forcing efficiencies those in Table S6.2 to S6.5? Can a reference be put in to these in the main text?

REPLY: OK done

CHANGES TO MANUSCRIPT: *we refer to the relevant tables in the SI in the respective sections 2.7.1 (aerosols)*

(P14 L22) The regional emission-normalized forcing SRs for aerosol precursors (in W m^{-2} Tg^{-1}) are given in Table S6.2 of the SI.

2.7.2 (indirect forcing)
(P15 L6) The global indirect forcing field associated with sulfate aerosols is shown in Fig. S6.1(d) of the SI an regional forcing SRs are listed in Table S6.2

and 2.7.3 (Radiative forcing by O3 and CH4)
(P15 L25) The details of the applied methodology for direct and indirect CH_4 forcing SRs are given in section S6.2 of the SI, including tables with the regional forcing efficiencies for all precursors (Tables S6.3 to S6.5).

And in the first line of section 2.7.4:
(P16 L2) The obtained emission-based forcing efficiencies (Tables S6.2 to S6.5 in the SI) are immediately useful for evaluating a set of short-lived climate pollutant climate metrics.

36) Section 3, P15, Lines 19 to 21 – Simplify point 1 to read better

REPLY: agree, we have rephrased the introduction of this section as follows

CHANGES TO MANUSCRIPT:

3 Results: validation of the reduced-form TM5-FASST

In this section we focus on the validation of regionally aggregated TM5-FASST_v0 outcomes (pollutant concentrations, exposure metrics, impacts), addressing specifically:

1 The additivity of individual pollutant responses as an approximation to obtain the response to combined precursor perturbations,
2 The linearity of the emission responses over perturbation ranges extending beyond the -20% perturbation
3 The FASST outcome versus TM5 for a set of global future emission scenarios that differ significantly from the reference scenario
4 FASST key-impact outcomes versus results from the literature for some selected case studies, with a focus on climate metrics, health impacts and crops.

37) Section 3.1, P16, Line 2 – reference is made to Annex 4 of the SI. Pleases clarify this reference as there is no Annex 4

REPLY: Indeed thanks for spotting.

CHANGES TO MANUSCRIPT: *reference is now correctly made to Table S3 (P17 L26)*

Additional standard -20% perturbation experiments P3 (NO_x only) and P5 (NO_x and NMVOC), as well as an additional set of perturbation simulations P1' to P5' over the range [-80%, +100%], listed in Table S3 of the SI, have been performed for a limited selection of representative source regions (Europe, USA, China, India, Japan) due to limited CPU resources.

38) Section 3.1.1, P16, Lines 12 to 14 – Is there a reason for the particular representative source regions selected in Table 2 e.g. South Africa for NOx

REPLY: In order to optimize computing time, NOx-only as well as the combined NOx-NMVOC perturbation regions were selected based on their presumed relevance in terms of impact, pace of expected emission changes in the future and geographical representativeness. South Africa was included as a case of rapidly developing economy in the Southern hemisphere and a possible case where it may be "safer" to explicitly calculate the NOx SR rather than applying gap filling.

39) Section 3.1.1, P16, Lines 19 to 22 – The explanation on these lines could be simplified

REPLY: done

CHANGES TO MANUSCRIPT: *we have rewritten the first part of section 3.1.1 as follows:*

3.1.1 Additivity and linearity of secondary inorganic $PM_{2.5}$ response:

Experiment P1, where BC, POM, SO_2 and NO_x emissions are simultaneously perturbed by -20% relative to base simulation P0, delivers SR matrices for primary components BC and POM, and a first-order approximation for the precursors SO_2 and NO_x whose emissions do not only affect SO_2 and NO_x gas concentrations but also lead to several secondary products (SO_2 forms ammonium sulfate, NO_x leads to O_3 and ammonium nitrate). Experiment P2 perturbs SO_2 only, while experiment P3 perturbs NO_x only (in this latter case, to limit the computational cost, computed for a limited set of representative source regions only).
We first test the hypothesis that the $PM_{2.5}$ response to the combined (NO_x + SO_2) -20% perturbation (P1) can be approximated by the sum of the single precursor perturbations responses (P2 + P3). Figure 2 summarizes the resulting change in SO_4^{2-}, NO_3^-, NH_4^+ and total inorganic $PM_{2.5}$ respectively for the selected source regionsFor Europe, the emission perturbations were applied over all European countries simultaneously, hence the responses are partly due to inter-regional transport from other countries. Following findings result from the perturbation experiments P1, P2 and P3:

(1) Sulfate shows a minor response to NO_x emissions, and likewise nitrate responds only slightly to SO_2 emissions and both perturbations are additive. In general the response is one order of magnitude lower than the direct formation of SO_4^{2-} and NO_3^- from SO_2 and NO_x respectively.(Fig. 2a, b).
(2) NH_4 responds to NO_x and SO_2 emissions with comparable magnitudes and in an additive way (Fig. 2c)
(3) A simultaneous -20% emission perturbation of SO_2 and NO_x behaves in an additive manner for what concerns the formation of secondary $PM_{2.5}$, i.e. the response of total sulfate, nitrate and ammonium to a combined NO_x and SO_2 perturbation can be approximated by the sum of the responses to the individual perturbations (Fig. 2d), i.e. P1 ≈ P2+P3. Scatterplots between P1 and P2+P3 for the regional averaged individual secondary products and total inorganic $PM_{2.5}$ are shown in Fig. S7.1of the SI

40) Section 3.1.1, P16, Lines 29 to 31 – Also there is a larger response to NO3 from increasing NOx emissions over India. Do you think that is this a particular issue for TM5 over India? Does this is cause issues for future prediction of NO3 aerosol from changes in NOx emissions over India?

REPLY: The reviewer correctly notices the large sensitivity of aerosol nitrate formation to NOx emissions in India. It is difficult to say whether this is a specific feature of TM5, or a more general feature of others models, as we are not aware of published sensitivity studies on NOx - aerosol NO3 in India. Moreover to our knowledge there are hardly any reliable NO3 observations available from India that could corroborate the calculated sensitivity. We will however highlight this feature in our paper, with a specific recommendation to devote more multi-model studies to this.

CHANGES TO MANUSCRIPT: *Modified / added following phrases*: *(P19 L2)*

The figure illustrates the general near-linear behaviour of regionally aggregated responses to single precursor emission perturbations for all regions, except for India where the linearity of the response to NO_x emissions breaks down for emission reductions beyond -50%. For India we further observe a relatively strong nitrate response to NOx emissions, with NO_3^- increasing by a factor of 3 for a doubling of NO_x emissions. We are not aware of reliable observations or other published NOx-aerosol sensitivity studies from that region that could corroborate the calculated

sensitivity. Because such a feature may strongly affect projected future $PM_{2.5}$ levels and associated impacts, we recommend devoting regional multi-model studies to this aspect.

41) Section 3.1.1, P17, Lines 8 to 11 – I don't think you can say that errors in the -80% case are larger than +100% for NOx. They look similar to me

REPLY: This part of the section has been rewritten to comply with earlier comments on statistic metrics

CHANGES TO MANUSCRIPT: .*We have rephrased the part of section3.1.1 dealing with linearity test under the large perturbations as follows: (P18 L30)*

Next we evaluate the hypothesis that the -20% perturbation responses can be extrapolated towards any perturbation range, as an approximation of a full TM5 simulation. Figure 3 shows, for the selected regions listed in Table S3 of the SI, the TM5 computed relative change in secondary $PM_{2.5}$ concentration versus the relative change in precursor emission in the range [-80%, +100%]. The figure illustrates the general near-linear behaviour of regionally aggregated responses to single precursor emission perturbations for all regions, except for India where the linearity of the response to NO_x emissions breaks down for emission reductions beyond -50%. For India we further observe a remarkably strong nitrate response to NO_x emissions, with NO_3^- increasing by a factor of 3 for a doubling of NO_x emissions, although the responses shown in Fig. 2 indicate that absolute changes (in $\mu g\ m^{-3}$) in NO_3 are relatively low and that secondary $PM_{2.5}$ in this region is dominated by SO_4. We are not aware of reliable observations or other published NO_x-aerosol sensitivity studies from that region that could corroborate this calculated sensitivity. Because such a feature may strongly affect projected future $PM_{2.5}$ levels and associated impacts, we recommend regional multi-model studies devote attention this feature Because the TM5-FASST linearization is based on the extrapolation of the -20% perturbation slope, concave-shaped trends in Fig. 3 indicate a tendency of TM5-FASST to over-predict secondary $PM_{2.5}$ at large negative or positive emission perturbations, and opposite for convex-shaped trends. Figure 4 illustrates the error introduced in regional secondary $PM_{2.5}$ concentrations responses when linearly extrapolating the regional -20% perturbation sensitivities to -80% (blue dots) and +100% (red dots) perturbations respectively. While the scatter plots for the single perturbations (Fig. 4 a,b,c) evaluate the linearity of the single responses, the panel showing the combined (SO_2+NO_x) perturbation (Fig. 4d) is a test for the linearity combined with additivity of SO_2 and NO_x perturbations over the considered range. In general, the linear approximation leads to a slight over-prediction of the resulting secondary $PM_{2.5}$ (i.e. the sum of sulfate, nitrate and ammonium) for all regions considered, in either perturbation direction. Table 4 shows regional statistical validation metrics (normalized mean bias NMB [%], mean bias MB [$\mu g\ m^{-3}$], and correlation coefficient, definitions are given in the Table Notes) for the grid-to-grid comparison between TM5-FASST and TM5-CTM of the response to the [-80%, 100%] perturbation simulations (with Europe presented as a single region). In terms of NMB, the FASST linearisation performs worst for the NO_x perturbations, with almost a factor 2 overestimate in Japan for an emission doubling. However, because of the already low NO_x emissions in this region, the absolute error (MB) remains below $0.2 \mu g\ m^{-3}$. In all considered perturbation cases, FASST shows a positive MB, except for the NO_x perturbation in India. In general, the highest NMB are observed for the regions where secondary $PM_{2.5}$ shows low response sensitivity to the applied perturbations and where the impact on the total $PM_{2.5}$ is therefore relatively low. Indeed, when considering the total resulting secondary $PM_{2.5}$ (i.e. the full right-hand side of Eq. 2, including the $PM_{2.5}$ base-concentration term containing primary and secondary components), regional averaged FASST secondary $PM_{2.5}$ values stay within 15% of TM5 (see Table S7.1of the SI). A break-down for the individual receptor regions within the European

zoom region of the linearisation error on the resulting total secondary $PM_{2.5}$ from individual and combined precursor perturbations is shown in Fig. S7.3 of the SI.

42) Section 3.1.2, P17, Lines 18 to 19 – Can you include references to back up the fact that combined NOx and NMVOCs emission perturbations will behave more linearly?

REPLY: We do not exactly say that combined and aligned NOx-VOC emission changes (in general) are behaving linearly, but, seen the fact that the ratio NOx/NMVOC determines the O_3 formation regime, combined emission changes of the same relative size and sign (in the way we applied them e.g. to establish the combined -20% perturbation responses) will not change the emission ratio and therefore preserve the O3 formation regime implying a linear behaviour. This is an implication of the statement made in the first phrase where we provide references.

CHANGES TO MANUSCRIPT*: we adapted the phrase as follows: (P20 L5)*

Because the NO_x/NMVOC ratio determines the O_3 response to emission changes, a perturbation with simultaneous NO_x and NMVOC emission changes of the same relative size is expected to behave more linearly than single perturbations since the chemical regime remains similar.

43) Section 3.1.2, P17, Line 31 – remove 'also here'

REPLY: done

44) Section 3.1.2, P17, Line 31 to 32 – Good agreement is found everywhere apart from China, Why?

REPLY: We presume the reviewer is referring to Fig. 5. Indeed for China the agreement between combined and sum of individual responses is – in absolute terms – slightly worse than most other regions, but in relative terms the sum of perturbations is within 10% of the combined one. We have added a scatterplot to Figure 5 to illustrate the over-all validity of additivity.
The underlying reason for the small deviations between combined and sum-of-individual responses has not been investigated in detail but, as stated above, is most probably linked to the fact that changing a single precursor emission strength changes the NOx/NMVOC ratio and could affect the O3 emission response regime.

CHANGES TO MANUSCRIPT: *P20 L19*

As shown in Fig. 5, for the -20% perturbations we find good agreement between the combined (NO_x + NMVOC) perturbation (open circles) with the sum of the individual precursor perturbation (black dots). This occurs even in situations where titration by NO causes a reverse response in O_3 concentration as is the case in most of Europe and the USA, indicating that a -20% perturbation in individual precursors appears not to change the prevailing O3 regime.

We also added a scatter plot to Fig. 5 to demonstrate the very good correspondence.

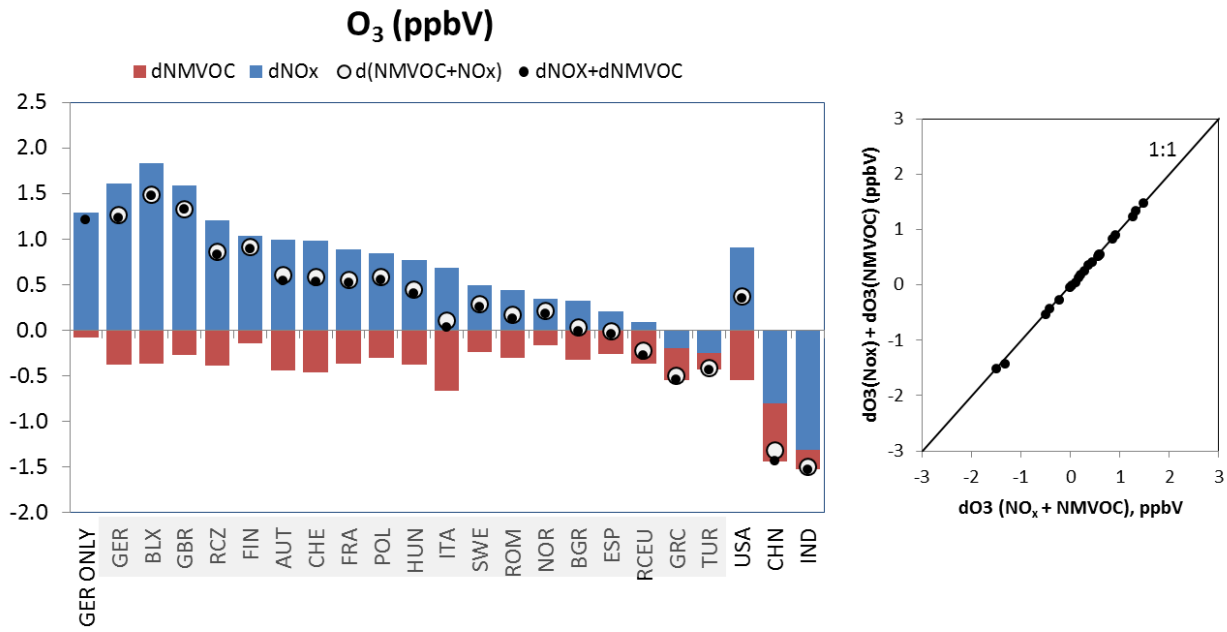

Figure 5: TM5-CTM response in annual mean population-weighted O_3 concentration (in ppbV) upon emitted precursor perturbation of -20% for selected source receptor regions. European regions were perturbed simultaneously. Red bar: response form NMVOC–only perturbation (simulation P4); blue bar: response form NO_x-only perturbation (simulation P3). Open circles: response from simultaneous (NMVOC + NO_x) perturbation (simulation P5). Black dots: sum of individual responses. Shaded regions are perturbed simultaneously as one European region. Right panel: scatter plot between O_3 response to combined and summed individual responses.

45) Section 3.1.2, P18, Line 2 – change 'Europa' to Europe

REPLY: done.

46) Section 3.1.2, P18, Lines 16 to 18 – If anything I would say FASST overestimates the change in TM5 (be it positive or negative) most of the time as the -80% points on the scatter plot tend to always above the 1:1 line (see major point 6 above).

REPLY: For a negative emission change, an origin-forced response slope below 1 (with points lying above the 1:1 line) indicates that the response between unperturbed and perturbed in FASST is lower than TM5, hence FASST underestimates the response upon an emission decrease and consequently overestimates the resulting concentration which is the sum of base and perturbation response (Eq. 2). A response slope larger than one for a positive emission change also corresponds to an over-prediction of the total concentration. We describe now more clearly in section 3.1 that we are evaluating the perturbation response (the change) and how an under/overestimation affects the total resulting concentration.

CHANGES TO MANUSCRIPT: *most of the section has been rewritten as follows: (P21 L6)*

Figure 7 illustrates the performance of the TM5-FASST approach versus TM5 for regional-mean annual mean ozone, health exposure metric 6mDMA1 (both evaluated as population-weighted mean), and for the crop-relevant exposure metrics AOT40 and M12 (both evaluated as area-weighted mean) over the extended emission perturbation range. In most cases the response (i.e. the *change* between base and perturbed case) to emission perturbations lies above the 1:1 line across the 4 metrics, indicating that FASST tends to over-predict the resulting metric (as a sum

of base concentration and perturbation). Of the four presented metrics, AOT40 is clearly the least robust one, which can be expected for a threshold-based metric that has been linearized. Tables 5 to 7 give the statistical metrics for the grid-to-grid comparison of the perturbation term between FASST and TM5 for the health exposure metric 6mDMA1, and crop exposure metrics AOT40 and M12 respectively. Statistical metrics for the total absolute concentrations (base concentration + perturbation term) are given in Tables S7.2 to S7.4 in the SI. As anticipated, the NO_x-only perturbation terms are showing the highest deviation, in particular for a doubling of emissions, however combined NO_x-NMVOC perturbations are reproduced fairly well for all regions, staying within 33% for a -80% perturbation for all 3 exposure metrics, and within 38% for an emission doubling for 6mDMA1 and M12, while the AOT40 metric is overestimated by 76 to 126% for emission doubling. The total resulting concentration over the entire perturbation range for single and combined NO_x and NMVOC perturbation agrees within 5% for 6mDMA1 and M12, and within 64% for AOT40. The mean bias is positive for both perturbations, for all metrics and over all analysed regions, except for crop metric M12 under a doubling of NMVOC emissions over Europe showing a small negative bias. The deviations for individual European receptor regions under single and combined NMVOC and NO_x perturbations for health and crop exposure metrics are shown in Figs. S7.4 to S7.6 of the SI.

47) Section 3.1.2, P18, Lines 21 to 22 – I am not sure that the linear fit is that good for the change in annual mean O3 in Figure 7a as there seems to be distinctive curvature in the +100% simulation for larger O3 reductions. I anticipate that this will be larger for certain months. The non-linear behaviour seems to occur to a lesser extent for other O3 metrics where a linear approximation is probably more justified. I think a change of wording for this statement is required to reflect the fact that a linear approximation does not represent the non-linear chemistry effects for large emission perturbation.

REPLY: The linear fits in Figure 7 were used as a guide to evaluate the overall correspondence of regional mean O3 metrics versus TM5, they are not the linear approximations used in FASST. (Each dot is obtained applying the region-specific SR coefficients for the respective precursors). Because this seems to cause confusion with the reader, we omitted the fittings and present the figure now only with the 1:1 line as a reference. Our statement refers to the observation that – except for AOT40 – the regional mean ozone metrics are relatively well represented by FASST (i.e. close to the 1:1 line) and in particular the FASST approximation reproduces the negative response to emission doubling (and positive response to emission reduction), typical for the titration regime.

CHANGES TO MANUSCRIPT: *We deleted the section using the slopes of the linear fits in Figs. 4 and 7– see also changes mentioned in previous comment*

48) Section 3.1.2, P18, Lines 25 to 28 – Check percentage numbers are correct as they don't appear to be the same as that shown on Figure S7.4 or in Table 3 e.g. -5 to 13% for M12 where on the Figure S7.4c I can't see anything below 0
49) Section 3.1.2, P18, Lines 28 to 30 – Same as above but for NMVOC
50) Section 3.1.2, P19, Line 1 – Same as above but for combined emission perturbation.

REPLY TO 47- 49:
The inconsistencies between values in the text and the figures were a consequence of a different statistical evaluation method, more in particular: the text vales were referring to the mean of all individual grid cell relative deviations, whereas the graphs were referring to the NMB as defined above (major comment 6). We report the values now consistently as NMB in text and figures.

51) Section 3.2, P19, Line 10 – remove 'e'

REPLY: done

52) Section 3.2, P19, Line 25 to 26 –In both scenarios emissions can change by >80% over some regions and precursors. The ability of FASST to predict such changes over regions needs to be highlighted in the results based on the breakdown of the linear approach for O3 at such high emission perturbations.

REPLY: Indeed a valid suggestion.

CHANGES TO MANUSCRIPT:

- *First of all, we became aware that the numbers reported in Table S8 (emission % changes relative to FASST reference) were wrong for all regions except Asia and Global – they have now been corrected (this does not affect the reported results)*
- *The introductory part of section 3.2 has been rewritten/rearranged mentioning some features of the scenario emissions, pointing to possible issues with combined emission changes that could not be addressed in the dedicated additivity/linearity simulations*
- *We have added new Figures, demonstrating that FASST does capture regional features both for low and high emission scenarios*

We modified the relevant paragraph to:

3.2 TM5-FASST_v0 versus TM5 for future emission scenarios

In this section we evaluate different combinations of precursor emission changes relative to the base scenario in a global framework. We take advantage of available TM5 simulations for a set of global emission scenarios which differ significantly in magnitude from the FASST base simulation, and as such provide a challenging test case to the application of the linear source-receptor relationships used in TM5-FASST. We assume that the full TM5 model provides valid evaluations of emission scenarios, and we test to what extent these simulations can be reproduced by the linear combinations of SRs implemented in the TM5-FASST_v0 model.

We use a set of selected policy scenarios prepared with the MESSAGE integrated assessment model in the frame of the Global Energy Assessment GEA (Rao et al., 2012, 2013; Riahi et al., 2012). These scenarios are the so called "frozen legislation" and "mitigation" emission variants for the year 2030 (named FLE2030, MIT2030 respectively), policy variants that describe two different policy assumptions on air pollution until 2030. These scenarios and there outcomes are described in detail in Rao et al. (2013), the scope of the present study is the inter-comparison between FASST and TM5 resulting pollutant concentration and exposure levels, as well as associated health impacts.

Major scenario features and emission characteristics are provided in section S8 of the SI. Table S8.1 shows the change in global emission strengths for the major precursors for both test scenarios, relative to the RCP2000 base, aggregated to the FASST 'master zoom' regions listed in Table S2.2. Emission changes for the selected scenarios mostly exceed the 20% emission perturbation amplitude from which the SRs were derived. Under the MIT2030 low emission scenario, all precursors and primary pollutants (except primary $PM_{2.5}$ in East-Asia and NH_3 in all regions) are showing a strong decrease compared to the RCP2000 reference scenario. The strongest decrease is seen in Europe (NO_x: -83%, SO_2: -93%, BC: -89%, primary $PM_{2.5}$ – 56%) while NH_3 is increasing by 14 to 46% across all regions. The FLE2030 scenario displays a global increase for all precursors, however with heterogeneous trends across regions. In Europe, North-America and Australia, the legislation in place, combined with use of less and cleaner fuels by 2030, leads to a decrease in pollutant emissions except for NH_3 and primary $PM_{2.5}$. On the other hand, very substantial emission increases are projected in East and South-East for BC, NO_x and primary $PM_{2.5}$. Anticipating possible linearity issues, we note that for both scenarios, in

all regions, SO_2 and NO_x emissions are evolving in the same direction, although not always with similar relative changes, while NH_3 is always increasing, which may induce linearity issues in the ammonium-sulfate-nitrate system. Regarding O_3 metrics, NMVOC and NO_x are evolving in the same direction, but also here we observe possible issues due to a changing emission ratio (in particular in Russia and Asia). We further note that not only the emission levels of these scenarios are different from the FASST base scenario (RCP year 2000), but also the spatial distribution of the emissions, at the resolution of grid cells, may differ from the reference set. We use FASST to compute $PM_{2.5}$ and ozone concentrations applying Eq. (2), i.e. considering the FLE2030 and MIT2030 emission scenarios as a perturbation on the FASST reference emission set (RCP year 2000).

The scope of TM5-FASST is to evaluate on a regional basis the impacts of policies that affect emissions of short-lived air pollutants and their precursors. Hence we average the resulting O_3 and $PM_{2.5}$ concentration and O_3 exposure metric 6mDMA1 over the each of the 56 FASST regions and compare them with the averaged TM5 results for the same regions.

Further, in a policy impact analysis framework, the *change* in pollutant concentrations between two scenarios (e.g. between a reference and policy case) is often more relevant than the absolute concentrations. We therefore present absolute concentrations as well as the change (delta) between the two GEA scenarios, evaluating the benefit of a mitigation scenario versus the frozen legislation scenario.

Figure 8 shows the FASST versus TM5 regional scatter plots for absolute and delta population-weighted mean anthropogenic $PM_{2.5}$ for all 56 FASST receptor regions while the population-weighted means over the 9 larger zoom areas are shown in Figure 9. Similarly annual mean population-weighted O_3 and 6mDMA1scatter plots are shown in Fig. 10, and the regional distribution in Fig. 11. The grid-cell statistics (mean, NMB, MB and R^2) over larger zoom areas are given in Tables 8 and 9 for $PM_{2.5}$ and 6mDMA1 respectively.

Figure 8 and Table 8 show that on a regional basis, the low emission scenario generally overestimates population-weighted $PM_{2.5}$ concentrations, with the highest negative bias in Europe and Asia, while the lowest deviation is found in Latin America and Africa. The agreement between FASST and TM5 is significantly better for the high emission scenario, in line with the findings in the previous section. As shown in Table 8, averaged over the larger zoom regions, we find that the relative deviation for $PM_{2.5}$ is within 11% for FLE2030, and within 28% for MIT2030, except for Europe where the (low) $PM_{2.5}$ concentration is overestimated by almost a factor of 2. The policy-relevant delta between the scenarios however is for all regions reproduced within 23%.

The ozone health metric 6mDMA1 is more scattered than annual mean ozone, and also here, as expected, the low emission scenario performs worse than the high emission one. Over larger zoom areas however the agreement is acceptable for both scenarios (FASST within 22% of TM5). Contrary to $PM_{2.5}$, the NMB for the delta 6mDMA1 between two scenarios is higher than the NMB on absolute concentrations, with a low bias for the delta metric of -38% and -45% for Europe and North-America respectively, and a high bias of 35 to 46% in Asia. However, the MB on the delta is of the same order or lower than the absolute concentrations (Table 9). This is a consequence of the fixed background ozone in the absolute concentration reducing the weight of the anthropogenic fraction in the relative error.

Figures 9 and 11 provide a general picture of the performance of FASST: despite the obvious uncertainties and errors introduced with the FASST linear approximation over larege emission changes compared to the RCP base run, at the level of regionally aggregated concentrations, a consistent result emerges both for absolute concentrations from the individual scenarios as for the policy-relevant delta.

And changed/added the following figures

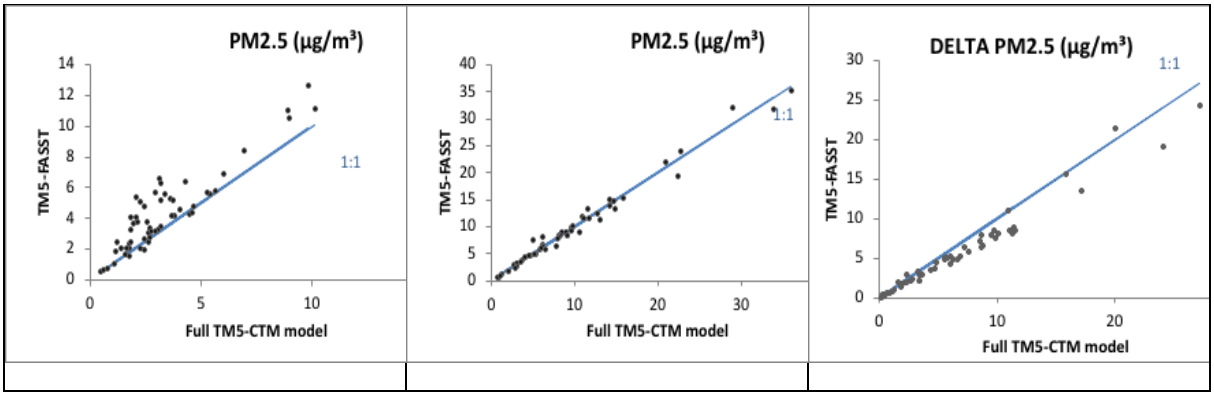

Figure 8: Population-weighted mean $PM_{2.5}$ concentration computed with TM5-FASST versus TM5-CTM for low emission scenarios MIT2030 (left), high emission scenario FLE2030 (middle) and the change between the two. Each point represents the population-weighted mean over a TM5-FASST receptor region. Blue line: 1:1 relation.

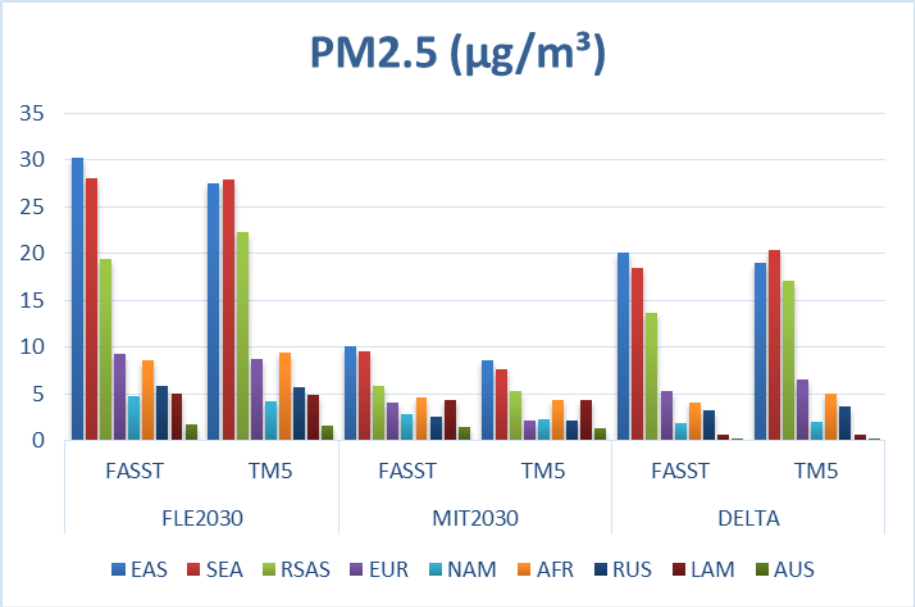

Figure 9 Total population-weighted anthropogenic $PM_{2.5}$ over larger FASST zoom areas, for the high (FLE2030) and low (MIT2030) emission scenarios, and the difference (delta) between both, computed with the full TM5 model and with FASST

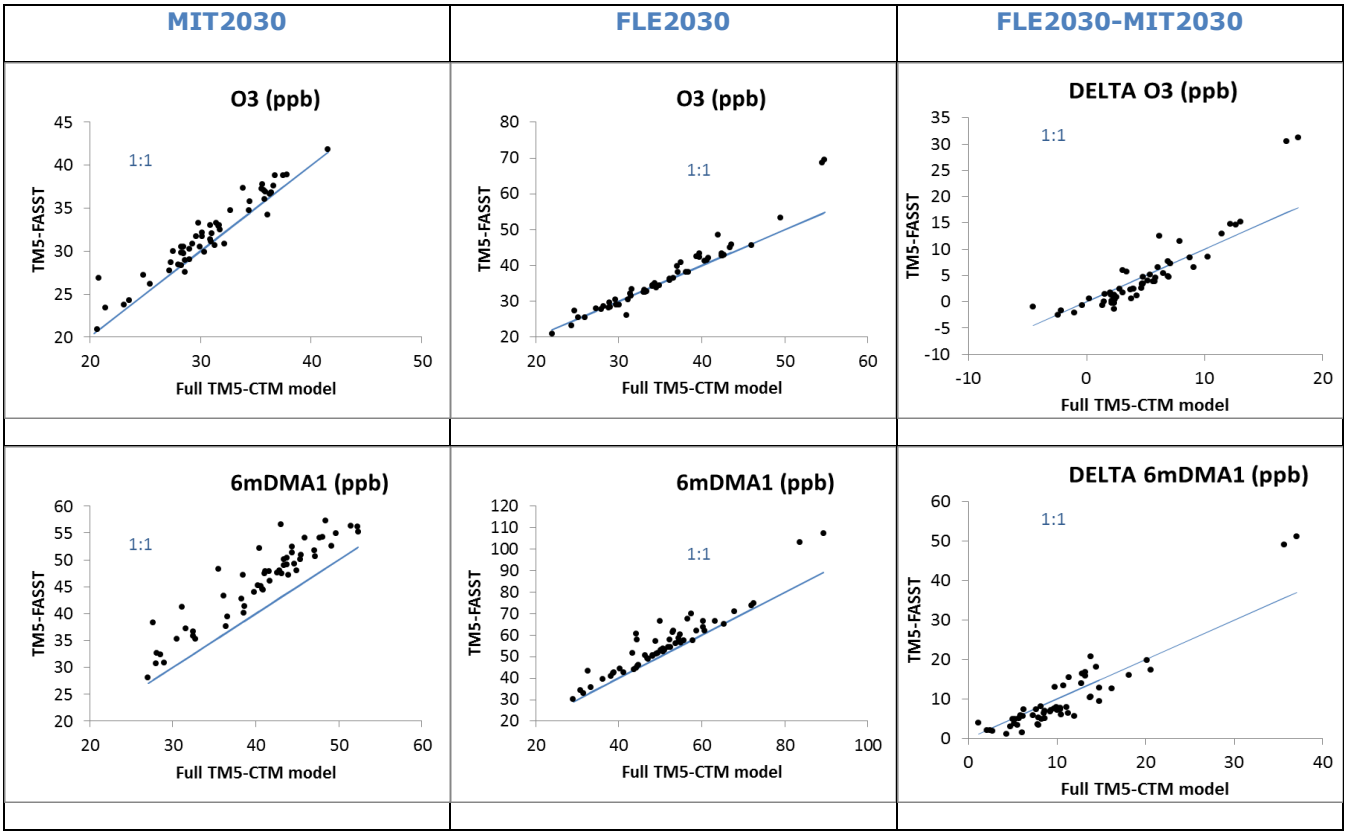

Figure 10: Population-weighted mean annual ozone (top) and ozone exposure metric 6mDMA1 (bottom) computed with TM5-FASST versus TM5-CTM for low emission scenarios MIT2030 (left), high emission scenario FLE2030 (middle) and the change between the two (right). Each point represents the population-weighted mean over a TM5-FASST receptor region. Blue line: 1:1 relation.

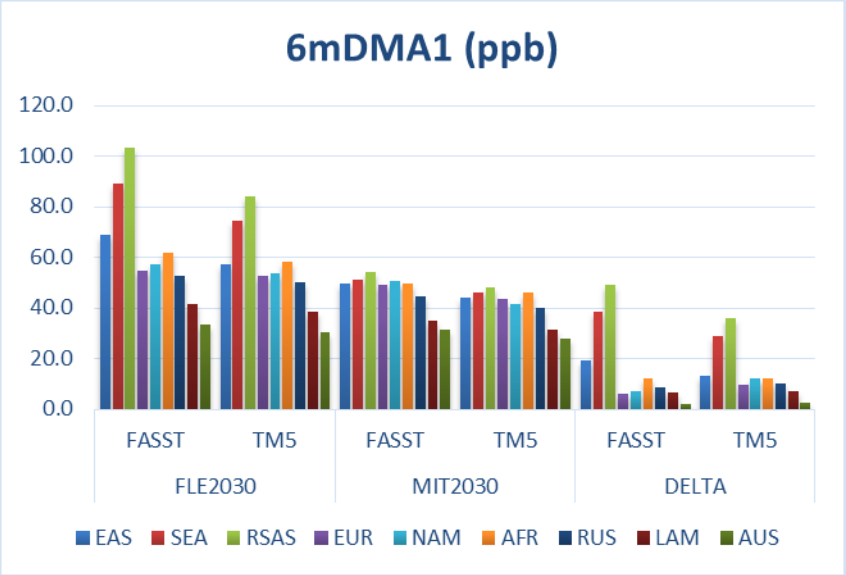

Figure 11: Total population-weighted anthropogenic $PM_{2.5}$ over larger FASST zoom areas, for the high (FLE2030) and low (MIT2030) emission scenarios, and the difference (delta) between both, computed with the full TM5 model and with FASST

53) Section 3.2, P20, Lines 3 to 5 – If it is more policy relevant to consider the change in pollutant concentrations between two scenarios than absolute concentrations, and FASST is a tool for the assessment of policy measures, then why is the difference not shown in place of absolute concentrations? Might be worth showing the change in concentrations in the main text and the absolute concentrations in the supplementary. Also it might be better to show the change between FLE and BASE, and MIT and BASE separately rather than the different between the two future scenarios

REPLY: We agree with the comment that from policy relevance perspective, putting more emphasis on the deltas makes sense. However, using in TM5FASST the RCP reference year 2000 as a common reference scenario is not very useful as here we are looking at a different scenario family (GEA) and a different year (2030). From policy perspective, comparing a 'policy' case (here: MIT2030) with a 'non-policy' case (here: FLE2030) for a given year immediately reveals the benefits of policy action. We therefore prefer to present the delta between the two GEA scenarios (with the additional benefit that this reduces the number of figures when showing the delta).

CHANGES TO MANUSCRIPT

As mentioned in the reply to the previous comment, we have rewritten most of the section. We include and discuss now both delta and totals for the two scenarios.

54) Section 3.2, P20, Line11 to 12 – I would say that FASST tends to underestimate the magnitude of change in TM5 for both annual mean and M6M O3, as most points are below the 1:1 line. (see major point 6 above).

REPLY: The referee made a correct observation; the slope is misleading here. This has been addressed with the changes made in text and the new figures (see previous comments)

55) Section 3.2, P20, Lines 15 to 21 – Only a very small discussion on the future evaluation of health metrics. Maybe expand slightly to include different regions and that FASST always overpredicts compared to TM5

REPLY: Agree, we have expanded the discussion of the intercomparison of the health impacts and included as well the delta in mortalities as from policy perspective this is relevant.

CHANGES TO MANUSCRIPT:

- *Added additional panels c to Figs. 10 and 11 showing the delta mortalities for* $PM_{2.5}$ *and O3 respectively*
- *Modified the health impact discussion as follows: (P23 L22)*

A major issue in air pollution or policy intervention impact assessments is the impact on human health; therefore we also evaluate the TM5-FASST outcome on air pollution premature mortalities with the TM5-based outcome, applying the same methodology on both TM5 and FASST outcomes. We evaluate mortalities from $PM_{2.5}$ using the IER functions (Burnett et al., 2014) and O_3 mortalities using the log-linear ER functions and RR's from Jerrett et al. (2009) respectively. Figure 12 ($PM_{2.5}$) and Fig. 13 (O_3) illustrate how FASST-computed mortalities compare to TM5, both as absolute numbers for each scenario, as well as the delta (i.e. the health benefit for MIT2030 relative to FLE2030). Regional differences in premature mortality numbers are mainly driven by population numbers. In line with the findings for the exposure metrics ($PM_{2.5}$ and 6mDMA1) FASST in general over-predicts the absolute mortality numbers, in particular in the low-emission case. For MIT2030, global $PM_{2.5}$ mortalities are overestimated by 19%, in Europe and North-America FASST even by 43%. In the FLE2030 case, we find a better agreement, with a global mortality over-prediction of 3% (for Europe and North-America 5% and 11% respectively). For the latter scenario, the highest deviation is found in Latin America (10 – 20%). O_3 mortalities are overestimated globally by 11% (7%) with regional agreement within 20% (14%) for MIT2030 (FLE2030). However, as shown by the error bars, the difference between FASST and TM5 is smaller than the uncertainty on the mortalities resulting from the uncertainty on RR's only. The potential health benefit of the mitigation versus the non-mitigation scenario (calculated as FLE2030 minus MIT2030 mortalities) is shown in Figs. 12c and 13c. Globally, FASST underestimates the reduction in global $PM_{2.5}$ mortalities by 17% with regional deviations ranging between -30% for Europe and North-America, and -12% for India. The global health benefit for ozone is underestimate by 2% for O_3, however as a net result of 11% overestimation in India and 12 to 59% underestimation in the other regions. The numbers corresponding to Figs. 12 and 13 are provided in Table S8.4 and S8.5 of the SI.
The error ranges presented here are obviously linked to the choice of the test scenarios and will for any particular scenario depend on the magnitude and the relative sign of the emission changes relative to RCP2000, but given the amplitude of the emission change for the currently two selected scenarios relative to RCP2000, these results support the usefulness of TM5-FASST as a tool for quick scenario screening.

56) Section 3.3.2, P22, Line 1 – relate discussion of text to labels on Figure 13 or define labels with more description in Figure 13 caption

REPLY: We presume this refers in particular to the labels in the b-panel (SLS M-O3 etc…). We have added the explanation in the figure caption.

CHANGES TO MANUSCRIPT*: changed the relevant section to*: *(P24 L2)*

Figure 15b shows the break-down by forcing component, including the direct contributions by aerosols, by short-lived precursors to O_3 (SLS S-O_3), their indirect effect on CH_4 (SLS I-CH_4) and associated long-term O_3 (SLS M-O_3), as well as CH_4 forcing from direct CH_4 emissions and its associated feedback on background ozone (CH3 O_3). Fig. 15a separates the contributions by emission sector..

And similar in the caption of Fig. 13 modified to:

Figure 13: Year 2000 radiative forcing from Unger et al. (2010), based on EDGAR year 2000 emissions and from TM5-FASST applied to RCP year 2000 (a) break-down by sector and by forcing component. Biomass burning includes both large scale fires and savannah burning; (b) total over all sectors. SLS S-O_3: direct contribution of short-lived species (SLS) to O_3; SLS I-CH_4: indirect contribution from SLS to CH_4; SLS M-O_3: indirect feedback from SLS on background ozone via the CH_4 feedback. CH_4 O_3: feedback of emitted CH_4 on background O_3

57) Section 3.3.3, P22, Line 13 – replace 'were' with 'where' and remove 'and'.

REPLY: done

58) Section 3.3.4, P22, Line 22 – remove 'implemented in FASST

REPLY: done

59) Section 3.3.4, P22, Line 24 – replace 'death cause' with 'a cause of death'.

REPLY: done

60) Section 3.3.4, P22, Line 25 to 27 – Could the difference in population and mortality rates between the two studies lead to some of the differences in Figure 14?

REPLY: This is unlikely for Figure 14 (now Figure 16) as it shows concentration changes, not mortalities. If the referee intends to refer to Fig. 15 (now 17), we mention in the text that Silva et al. use indeed different population and base mortality projections. In particular – as mentioned - the projection for respiratory base mortality rates (which is relevant only for the O3 health impact and not for PM2.5) for 2050 in Silva et al. is very different from the values used in FASST (where they are constant compared to 2030 base mortality rates). The discrete dots in the O3 mortality graph are a simple attempt to demonstrate the impact in FASST of using these different mortality rates.

61) Section 3.3.4, P23, Line 9 to 11 – How can FASST account for inter-model variability in its results? I think that this is mentioned as a future development so needs to be linked to that here.

REPLY: We intend to say that the difference between FASST and the ACCMIP model ensemble for what concerns O3, is probably not due to a poor performance of FASST (which is a fairly good approximation of TM5) but rather a consequence of generally occurring differences between models.

CHANGES TO MANUSCRIPT: *modified the phrase as follows: (P28 L7)*

The ozone exposure metric 6mDMA1 falls within the range of the ACCMIP model ensemble for 2030 - 2050, but the slope between 2030 and 2050 is lower than for the ACCMIP ensemble mean, i.e. FASST shows a much lower response sensitivity for O_3 to changing emissions between 2030 and 2050 than the ACCMIP models (-1ppb from 2030 to 2050 in FASST, versus -3ppb for the ACCMIP mean). Given our previous observation that FASST reproduces TM5 relatively well, this indicates that inter-model variability is a stronger factor in the model uncertainty than the reduced-form approach.

62) Section 3.3.4, P23, Line 17 – replace 'While also' with 'Whilst calculated'

REPLY: It seems the use of "while" or "whilst" is interchangeable in English language. As non-native English speaker it feels more comfortable to use "while".

63) Section 3.3.4, P23, Line 18 to 21 – Why does the different baseline mortality and population statistics have such a big impact on O3 mortality rates but not PM2.5?

REPLY: The reason is that respiratory mortality is not considered a cause of death from PM2.5; the GBD methodology includes COPD, LC, IHD and Stroke for $PM_{2.5}$ and respiratory disease for O3.

CHANGES TO MANUSCRIPT: *added the following phrase: (P28 L25)*

Respiratory mortality is not considered as a cause of death for PM2.5, which explains why a similar disagreement is not observed in the $PM_{2.5}$ mortality trend in Fig. 17b.

64) Section 3.3.4, P23, Line 27 to 31 – Could a little more discussion on regional mortality burdens be put into the main text. Interesting differences between regions and for RCP2.6 vs RCP8.5.

REPLY: Although the scope of this paper is not to make a scenario analysis or assess trends and impacts across regions, but rather to validate the FASST model, we agree that some more discussion is useful.

CHANGES TO MANUSCRIPT*: paragraph has been rewritten as follows: (P28 L28)*

A regional break-down of mortality burden from $PM_{2.5}$ in 2030 and 2050, relative to exposure to year 2000 concentrations, for major world regions and for the globe is shown in Figures S9.1 and S9.2 of the SI. Compared to Fig. 17 which shows the global mortality trends as a combined effect of changing population, mortality rates and pollution level, here the effect of changing population and baseline mortality is eliminated by exposing the evaluated year's population to pollutant levels of the relevant year and to RCP year 2000 levels respectively, and calculating the change between the two resulting mortality numbers. FASST reproduces the over-all observed trends across the regions: we see substantial reductions in North America and Europe in 2030, while in East Asia significant improvements in air quality impacts are realized after 2030. For the India region, all scenarios project a worsening of the situation. The global trend is dominated by the changes in East Asia. The observed differences between FASST and ACCMIP ensemble are not insignificant and partly due to different mortality and population statistics in particular for the year 2050, still they are consistent with the findings in the previous section: FASST tends to overestimate absolute $PM_{2.5}$ concentrations for emission scenarios different from RCP2000, and consequently tends to under-predict the benefit of emission reductions, while over-predicting the impact of increasing emissions.

65) Section 4, P24, Line 17 to 18 – Make statement in this sentence less strong by inserting 'tend to' between 'metrics' and 'remain'.

REPLY: done

66) Section 4, P24, Lines 21 to 23 – I think the first two sentences could be re-written to simply specify that because the emissions and meteorology are fixed the source receptor matrices remain fixed. Also I think the work 'arbitrary' should be removed.

REPLY: we have rephrased the sentences

CHANGES TO MANUSCRIPT: *(P30 L22)*

Another issue for caution relates to the FASST analysis of emission scenarios with spatial distribution that differs from the FASST reference scenario (RCP year 2000). The definition of the source regions when establishing the SR matrices implicitly freezes the spatial distribution of pollutant emissions within each region, and therefore the reduced-form model cannot deal with intra-regional spatial shifts in emissions.

67) Section 4, P24, Lines 25 – remove repetitive statement of 'compared to the base simulation year 2000'.

REPLY: done

68) Section 4, P24, Lines 27 – remove 'be

REPLY: done

69) Section 4, P24, Lines 30 to 31 – reword sentence to 'It can be expected that errors will be larger for the newer generation scenarios with dynamic allocation of emission across countries and macro-regions'

REPLY: done

70) Section 4, P25, Lines 5 to 7 – Sectors are mentioned that can't be assessed but little has been mentioned about shipping and aviation which can be assessed and are specifically included as a source region in FASST. I think it is worth mentioning these source regions in this section

REPLY: Thank you for bringing this up – indeed worth mentioning.

CHANGES TO MANUSCRIPT: *added in discussion P31 L7:*

This limitation however does not apply to international shipping and aviation for which specific SR matrices have been established.

71) Section 5, P25, Line 32 – removal of '....' at end of page

REPLY: done

72) Section 5, P26, Line 6 – subscripts for O3 and $PM_{2.5}$ required

REPLY: done

73) Section 5, P26, Line 19 – Slightly more detail could be provided on how the HTAP2 modelling exercise will inform/improve TM5-FASST, especially as TM5 was not a model that participated in HTAP2.

REPLY: The FASST architecture makes it possible to include new or additional SR matrices, even when they have been obtained from different models and with different regional definitions. SR simulations are now available from various models participating in HTAP2, but the 'required' and 'desired' simulations have not been fully completed by all participating models, and gapfilling method has been proposed (Turnock et al., 2018). Therefore a tool like FASST which

could bring this knowledge in a common structure, synthesizing the available data in an ensemble approach and make it accessible and applicable for interested users, would create a great added value. In the context of the UNECE/CLRTAP TF HTAP such a tool is already under development.

CHANGES TO MANUSCRIPT: *added the paragraph P34 L22*

The FASST architecture allows for an implementation of new or additional SR matrices, for instance new HTAP2 model ensemble mean matrices, each one accompanied by an ensemble standard deviation matrix to include the model variability in the results. Efforts are now underway to create a new web-based and user-friendly HTAP-FASST version, operating under the same principles as TM5-FASST, but based on an up-to-date reference simulation and underlying meteorology, thus creating a link between the knowledge generated by the HTAP scientific community and interested policy-oriented users. Indeed, similar to how the development of TM5-FASST was built upon extending the HTAP1 experiments in a single model context, the regional definitions and sector definitions used in HTAP2 (Galmarini et al., 2017; Koffi et al., 2016) were largely synchronized with the TM5-FASST set-up, increasing the community's capacity for multi-model assessments of hemispheric pollution. It is intended that the lessons-learned are informing the HTAP2 exercise

74) Figure 14 – I find that the grey lines mask out the black lines in some instances and I think the Figure would look better if the grey lines could be made less bold or more transparent. Also I am not sure why there is a different number of grey lines on each part of the Figure. Did a different number of models submit results for each experiment?

REPLY: Indeed, in ACCMIP not all models participated in each experiment, hence the different numbers. We have modified the figure to make the black lines more visible, and added information to the legenda.

CHANGES TO MANUSCRIPT: *modified Fig. 14 (now 16) into*

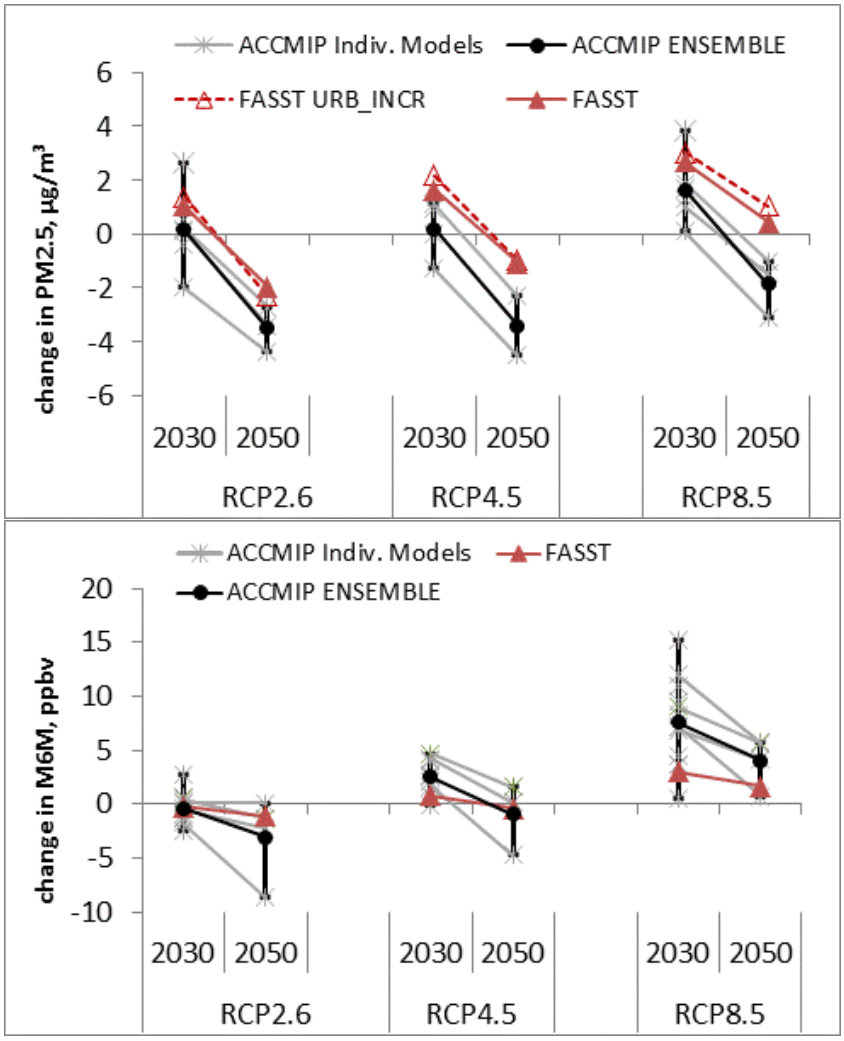

Figure 16: Global population-weighted differences (scenario year minus year 2000) (a) in annual mean $PM_{2.5}$ concentrations and (b) in O_3 exposure metric 6mDMA1 for 3 RCP scenarios in each future year, from the ACCMIP model ensemble (Silva et al., 2016) (black symbols and lines) and TM5-FASST_v0 (red symbols and lines). FASST URB_INCR: including the urban increment correction. Grey symbols: results from individual ACCMIP models. Grey lines connect results from a single model. Not all models have provided data for all scenarios. ACCMIP error bars represent the range (min, max) across the ACCMIP ensemble.

75) Table S3 – Certain lines in the table seem to be missing any information. e.g. P5 Germany, P4 USA, P5 Japan.

REPLY: That's a correct observation, in fact in those cases the experiments were not performed.

CHANGES TO MANUSCRIPT: *we removed the irrelevant lines, added a prime to the Pi' to distinguish from the -20% perturbations and added a line for the additional P1' simulations that were performed as well.*

76) Figure S3.3 – Why has the sign been reversed? For a 20% reduction in CH4 you would expect a decrease in O3 concentrations but the figure shows positive changes. This seems confusing

REPLY: Apologies for the confusion. The SR response field were stored as a positive change to a positive perturbation (although the perturbation runs were performed as negative perturbations resulting in a negative response)..

CHANGES TO MANUSCRIPT*: The caption has now been modified to:*

Figure S3.3 Decrease in annual mean surface O_3 for a 20% decrease in year 2000 CH4 concentration, i.e. 1760 to 1408 ppb (TF-HTAP1 SR1-SR2 scenarios)

77) Section S4.1, Equation 4.4 – I am not sure I can follow how the INCR formulation was derived and why it includes the (fup)2 terms.

REPLY: we added one intermediate step in the calculation that explains how the quadratic terms in fup are obtained.

CHANGES TO MANUSCRIPT*: added*

The population-weighted concentration is calculated as

$$C_{BC,TM5}^{pop} = f_{up}C_{BC,URB} + (1 - f_{up})C_{BC,RUR} \quad [4.3]$$

78) Figure S5 – hard to decipher the different lines on the graph. Cannot see red lines most of the time. Please make clearer

REPLY: we have decreased the size of the dots and increased the line width so it is better visible

CHANGES TO MANUSCRIPT: *figures modified in the folowing way:*

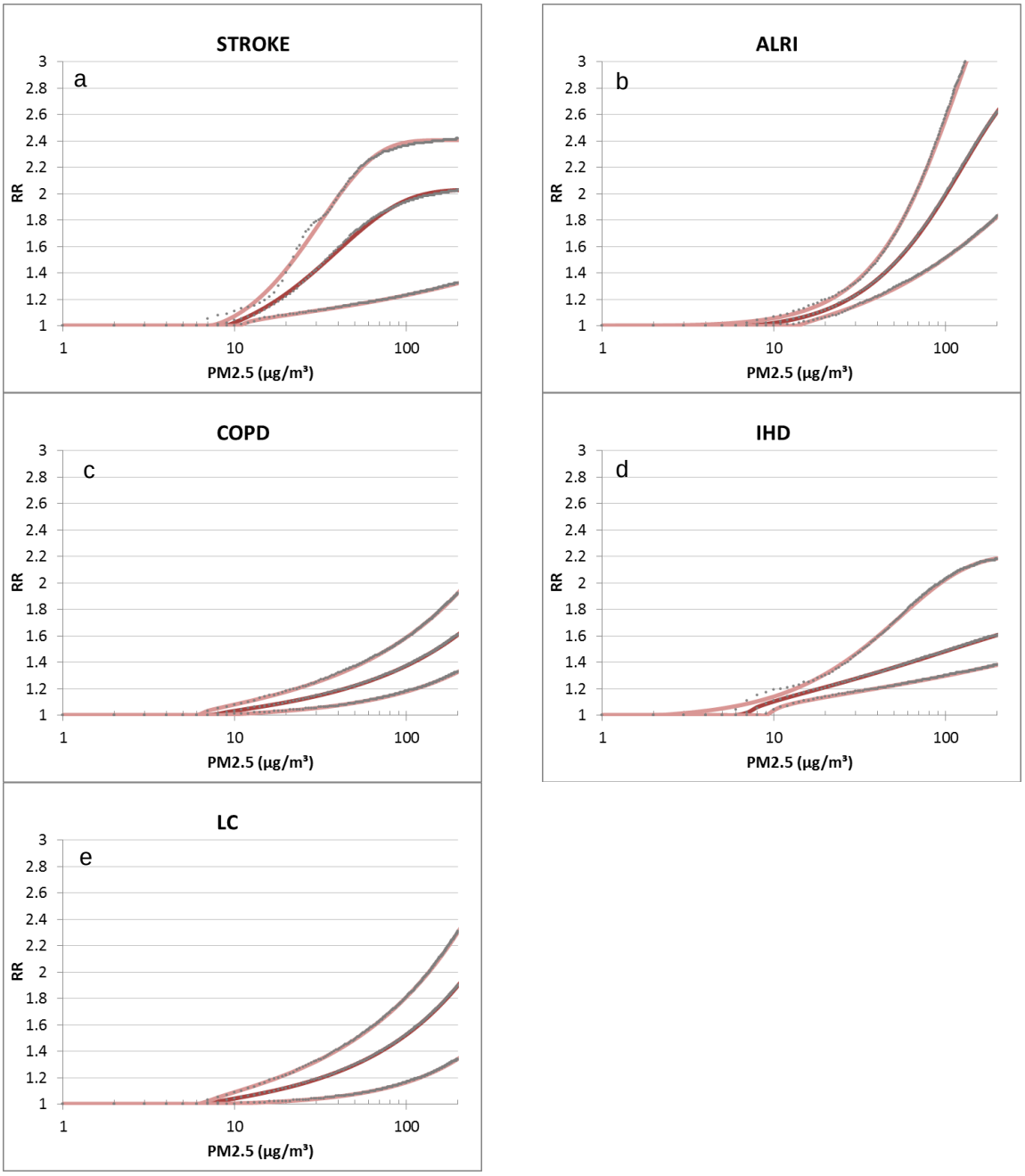

Figure S5.1: Dashed line: Median and 95% CI of the relative risk (RR) as a function of exposure to $PM_{2.5}$ from 1000 Monte Carlo samples provided by Burnett et al. (2014). Red lines: fitted curves for all-age IER functions for 5 mortality causes, using the parameters listed in Table S6.1 (this work). (a): Stroke, (b): Acute Lower Respiratory Airways Infections (c) Chronic Obstructive Pulmonary Disease (d) Ischaemic Heart Disease (e) Lung Cancer

79) Section S6.1, P24, Line 166 – 'Table S7.1' should be Table S6.1.

REPLY: OK done

References:

Andersson, C., Langner, J. and Bergström, R.: Interannual variation and trends in air pollution over Europe due to climate variability during 1958–2001 simulated with a regional CTM coupled to the ERA40 reanalysis, Tellus B, 59(1), 77–98, doi:10.1111/j.1600-0889.2006.00196.x, 2007.

Avnery, S., Mauzerall, D. L., Liu, J. and Horowitz, L. W.: Global crop yield reductions due to surface ozone exposure: 1. Year 2000 crop production losses and economic damage, Atmos. Environ., 45(13), 2284–2296, doi:10.1016/j.atmosenv.2010.11.045, 2011.

Boucher, O. and Lohmann, U.: The sulfate-CCN-cloud albedo effect: a sensitivity study with two general circulation models, Tellus Ser. B, 47 B(3), 281–300, 1995.

Burnett, R. T., Pope, C. A., III, Ezzati, M., Olives, C., Lim, S. S., Mehta, S., Shin, H. H., Singh, G., Hubbell, B., Brauer, M., Anderson, H. R., Smith, K. R., Balmes, J. R., Bruce, N. G., Kan, H., Laden, F., Prüss-Ustün, A., Turner, M. C., Gapstur, S. M., Diver, W. R. and Cohen, A.: An Integrated Risk Function for Estimating the Global Burden of Disease Attributable to Ambient Fine Particulate Matter Exposure, Environ. Health Perspect., doi:10.1289/ehp.1307049, 2014.

Cohen, A. J., Brauer, M., Burnett, R., Anderson, H. R., Frostad, J., Estep, K., Balakrishnan, K., Brunekreef, B., Dandona, L., Dandona, R., Feigin, V., Freedman, G., Hubbell, B., Jobling, A., Kan, H., Knibbs, L., Liu, Y., Martin, R., Morawska, L., Pope, C. A., III, Shin, H., Straif, K., Shaddick, G., Thomas, M., van Dingenen, R., van Donkelaar, A., Vos, T., Murray, C. J. L. and Forouzanfar, M. H.: Estimates and 25-year trends of the global burden of disease attributable to ambient air pollution: an analysis of data from the Global Burden of Diseases Study 2015, The Lancet, 389(10082), 1907–1918, doi:10.1016/S0140-6736(17)30505-6, 2017.

Dentener, F., Stevenson, D., Cofala, J., Mechler, R., Amann, M., Bergamaschi, P., Raes, F. and Derwent, R.: The impact of air pollutant and methane emission controls on tropospheric ozone and radiative forcing: CTM calculations for the period 1990-2030, Atmos Chem Phys, 5(7), 1731–1755, doi:10.5194/acp-5-1731-2005, 2005.

Dentener, F., Kinne, S., Bond, T., Boucher, O., Cofala, J., Generoso, S., Ginoux, P., Gong, S., Hoelzemann, J. J., Ito, A., Marelli, L., Penner, J. E., Putaud, J.-P., Textor, C., Schulz, M., Van, D. W. and Wilson, J.: Emissions of primary aerosol and precursor gases in the years 2000 and 1750 prescribed data-sets for AeroCom, Atmospheric Chem. Phys., 6(12), 4321–4344, 2006.

Dentener, F., Keating, T., Akimoto, H., Pirrone, N., Dutchak, S., Zuber, A., Convention on Long-range Transboundary Air Pollution, United Nations and UNECE Task Force on Emission Inventories and Projections, Eds.: Hemispheric transport of air pollution 2010: prepared by the Task Force on Hemispheric Transport of Air Pollution acting within the framework of the Convention on Long-range Transboundary Air Pollution, United Nations, New York ; Geneva., 2010.

Edwards, J. M. and Slingo, A.: Studies with a flexible new radiation code. I: Choosing a configuration for a large-scale model, Q. J. R. Meteorol. Soc., 122(531), 689–719, 1996.

Farina, S. C., Adams, P. J. and Pandis, S. N.: Modeling global secondary organic aerosol formation and processing with the volatility basis set: Implications for anthropogenic secondary organic aerosol, J. Geophys. Res. Atmospheres, 115(D9), doi:10.1029/2009JD013046, 2010.

Fenech, S., Doherty, R. M., Heaviside, C., Vardoulakis, S., Macintyre, H. L. and O'Connor, F. M.: The influence of model spatial resolution on simulated ozone and fine particulate matter for Europe: implications for health impact assessments, Atmospheric Chem. Phys., 18(8), 5765–5784, doi:10.5194/acp-18-5765-2018, 2018.

Fiore, A. M., Jacob, D. J., Field, B. D., Streets, D. G., Fernandes, S. D. and Jang, C.: Linking ozone pollution and climate change: The case for controlling methane, Geophys. Res. Lett., 29(19), 25–1, 2002.

Fiore, A. M., West, J. J., Horowitz, L. W., Naik, V. and Schwarzkopf, M. D.: Characterizing the tropospheric ozone response to methane emission controls and the benefits to climate and air quality, J. Geophys. Res. Atmospheres, 113(8), doi:10.1029/2007JD009162, 2008.

Fiore, A. M., Dentener, F. J., Wild, O., Cuvelier, C., Schultz, M. G., Hess, P., Textor, C., Schulz, M., Doherty, R. M., Horowitz, L. W., MacKenzie, I. A., Sanderson, M. G., Shindell, D. T., Stevenson, D. S., Szopa, S., Van Dingenen, R., Zeng, G., Atherton, C., Bergmann, D., Bey, I., Carmichael, G., Collins, W. J., Duncan, B. N., Faluvegi, G., Folberth, G., Gauss, M., Gong, S., Hauglustaine, D., Holloway, T., Isaksen, I. S. A., Jacob, D. J., Jonson, J. E., Kaminski, J. W., Keating, T. J., Lupu, A., Marmer, E., Montanaro, V., Park, R. J., Pitari, G., Pringle, K. J., Pyle, J. A., Schroeder, S., Vivanco, M. G., Wind, P., Wojcik, G., Wu, S. and Zuber, A.: Multimodel estimates of intercontinental source-receptor relationships for ozone pollution, J. Geophys. Res. Atmospheres, 114(D4), doi:10.1029/2008JD010816, 2009.

Forouzanfar, M. H., Afshin, A., Alexander, L. T., Biryukov, S., Brauer, M., Cercy, K., Charlson, F. J., Cohen, A. J., Dandona, L., Estep, K., Ferrari, A. J., Frostad, J. J., Fullman, N., Godwin, W. W., Griswold, M., Hay, S. I., Kyu, H. H., Larson, H. J., Lim, S. S., Liu, P. Y., Lopez, A. D., Lozano, R., Marczak, L., Mokdad, A. H., Moradi-Lakeh, M., Naghavi, M., Reitsma, M. B., Roth, G. A., Sur, P. J., Vos, T., Wagner, J. A., Wang, H., Zhao, Y., Zhou, M., Barber, R. M., Bell, B., Blore, J. D., Casey, D. C., Coates, M. M., Cooperrider, K., Cornaby, L., Dicker, D., Erskine, H. E., Fleming, T., Foreman, K., Gakidou, E., Haagsma, J. A., Johnson, C. O., Kemmer, L., Ku, T., Leung, J., Masiye, F., Millear, A., Mirarefin, M., Misganaw, A., Mullany, E., Mumford, J. E., Ng, M., Olsen, H., Rao, P., Reinig, N., Roman, Y., Sandar, L., Santomauro, D. F., Slepak, E. L., Sorensen, R. J. D., Thomas, B. A., Vollset, S. E., Whiteford, H. A., Zipkin, B., Murray, C. J. L., Mock, C. N., Anderson, B. O., Futran, N. D., Anderson, H. R., Bhutta, Z. A., Nisar, M. I., Akseer, N., Krueger, H., Gotay, C. C., Kissoon, N., Kopec, J. A., Pourmalek, F., Burnett, R., Abajobir, A. A., Knibbs, L. D., Veerman, J. L., Lalloo, R., Scott, J. G., Alam, N. K. M., Gouda, H. N., Guo, Y., McGrath, J. J., Charlson, F. J., Erskine, H. E., Jeemon, P., Dandona, R., Goenka, S., Kumar, G. A., et al.: Global, regional, and national comparative risk assessment of 79 behavioural, environmental and occupational, and metabolic risks or clusters of risks, 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015, The Lancet, 388(10053), 1659–1724, doi:10.1016/S0140-6736(16)31679-8, 2016.

Galmarini, S., Koffi, B., Solazzo, E., Keating, T., Hogrefe, C., Schulz, M., Benedictow, A., Jurgen, G., Janssens-Maenhout, G., Carmichael, G., Fu, J. and Dentener, F.: Technical note: Coordination and harmonization of the multi-scale, multi-model activities HTAP2, AQMEII3, and MICS-Asia3: Simulations, emission inventories, boundary conditions, and model output formats, Atmospheric Chem. Phys., 17(2), 1543–1555, doi:10.5194/acp-17-1543-2017, 2017.

IIASA and FAO: Global Agro-Ecological Zones V3.0, [online] Available from: http://www.gaez.iiasa.ac.at/ (Accessed 11 November 2016), 2012.

Jerrett, M., Burnett, R. T., Arden, P. I., Ito, K., Thurston, G., Krewski, D., Shi, Y., Calle, E. and Thun, M.: Long-term ozone exposure and mortality, N. Engl. J. Med., 360(11), 1085–1095, doi:10.1056/NEJMoa0803894, 2009.

Kanakidou, M., Seinfeld, J. H., Pandis, S. N., Barnes, I., Dentener, F. J., Facchini, M. C., Dingenen, R. V., Ervens, B., Nenes, A., Nielsen, C. J., Swietlicki, E., Putaud, J. P., Balkanski, Y., Fuzzi, S., Horth, J., Moortgat, G. K., Winterhalter, R., Myhre, C. E. L., Tsigaridis, K., Vignati, E., Stephanou, E. G. and Wilson, J.: Organic aerosol and global climate modelling: a review, Atmospheric Chem. Phys., 5(4), 1053–1123, doi:10.5194/acp-5-1053-2005, 2005.

Kirschke, S., Bousquet, P., Ciais, P., Saunois, M., Canadell, J. G., Dlugokencky, E. J., Bergamaschi, P., Bergmann, D., Blake, D. R., Bruhwiler, L., Cameron-Smith, P., Castaldi, S., Chevallier, F.,

Feng, L., Fraser, A., Heimann, M., Hodson, E. L., Houweling, S., Josse, B., Fraser, P. J., Krummel, P. B., Lamarque, J.-F., Langenfelds, R. L., Le Quéré, C., Naik, V., O'Doherty, S., Palmer, P. I., Pison, I., Plummer, D., Poulter, B., Prinn, R. G., Rigby, M., Ringeval, B., Santini, M., Schmidt, M., Shindell, D. T., Simpson, I. J., Spahni, R., Steele, L. P., Strode, S. A., Sudo, K., Szopa, S., van der Werf, G. R., Voulgarakis, A., van Weele, M., Weiss, R. F., Williams, J. E. and Zeng, G.: Three decades of global methane sources and sinks, Nat. Geosci., 6(10), 813–823, doi:10.1038/ngeo1955, 2013.

Koffi, B., Dentener, F., Janssens-Maenhout, G., Guizzardi, D., Crippa, M., Diehl, T., Galmarini, S. and Solazzo, E.: Hemispheric Transport Air Pollution (HTAP): Specification of the HTAP2 experiments, Publications Office of the European Union, Luxembourg. [online] Available from: http://publications.jrc.ec.europa.eu/repository/bitstream/JRC102552/lbna28255enn.pdf (Accessed 14 December 2017), 2016.

Krewski, D., Jerrett, M., Burnett, R. T., Ma, R., Hughes, E. and Shi, Y.: Extended Follow-Up and Spatial Analysis of the American Cancer Society Study Linking Particulate Air Pollution and Mortality., Research Report, Health Effects Institute, Boston., 2009.

Li, Y., Henze, D. K., Jack, D. and Kinney, P. L.: The influence of air quality model resolution on health impact assessment for fine particulate matter and its components, Air Qual. Atmosphere Health, 9(1), 51–68, doi:10.1007/s11869-015-0321-z, 2016.

Lim, S. S., Vos, T., Flaxman, A. D., Danaei, G., Shibuya, K., Adair-Rohani, H., Amann, M., Anderson, H. R., Andrews, K. G., Aryee, M., Atkinson, C., Bacchus, L. J., Bahalim, A. N., Balakrishnan, K., Balmes, J., Barker-Collo, S., Baxter, A., Bell, M. L., Blore, J. D., Blyth, F., Bonner, C., Borges, G., Bourne, R., Boussinesq, M., Brauer, M., Brooks, P., Bruce, N. G., Brunekreef, B., Bryan-Hancock, C., Bucello, C., Buchbinder, R., Bull, F., Burnett, R. T., Byers, T. E., Calabria, B., Carapetis, J., Carnahan, E., Chafe, Z., Charlson, F., Chen, H., Chen, J. S., Cheng, A. T.-A., Child, J. C., Cohen, A., Colson, K. E., Cowie, B. C., Darby, S., Darling, S., Davis, A., Degenhardt, L., Dentener, F., Des Jarlais, D. C., Devries, K., Dherani, M., Ding, E. L., Dorsey, E. R., Driscoll, T., Edmond, K., Ali, S. E., Engell, R. E., Erwin, P. J., Fahimi, S., Falder, G., Farzadfar, F., Ferrari, A., Finucane, M. M., Flaxman, S., Fowkes, F. G. R., Freedman, G., Freeman, M. K., Gakidou, E., Ghosh, S., Giovannucci, E., Gmel, G., Graham, K., Grainger, R., Grant, B., Gunnell, D., Gutierrez, H. R., Hall, W., Hoek, H. W., Hogan, A., Hosgood III, H. D., Hoy, D., Hu, H., Hubbell, B. J., Hutchings, S. J., Ibeanusi, S. E., Jacklyn, G. L., Jasrasaria, R., Jonas, J. B., Kan, H., Kanis, J. A., Kassebaum, N., Kawakami, N., Khang, Y.-H., Khatibzadeh, S., Khoo, J.-P., Kok, C., et al.: A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990-2010: A systematic analysis for the Global Burden of Disease Study 2010, The Lancet, 380(9859), 2224–2260, doi:10.1016/S0140-6736(12)61766-8, 2012.

Liu, X., Penner, J. E., Das, B., Bergmann, D., Rodriguez, J. M., Strahan, S., Wang, M. and Feng, Y.: Uncertainties in global aerosol simulations: Assessment using three meteorological data sets, J. Geophys. Res., 112(D11), doi:10.1029/2006JD008216, 2007.

Malley, C. S., Henze, D. K., Kuylenstierna, J. C. I., Vallack, H., Davila, Y., Anenberg, S. C., Turner, M. C. and Ashmore, M.: Updated Global Estimates of Respiratory Mortality in Adults ≥30Years of Age Attributable to Long-Term Ozone Exposure., Environ. Health Perspect., 125(8), 087021, doi:10.1289/EHP1390, 2017.

Mills, G., Buse, A., Gimeno, B., Bermejo, V., Holland, M., Emberson, L. and Pleijel, H.: A synthesis of AOT40-based response functions and critical levels of ozone for agricultural and horticultural crops, Atmos. Environ., 41(12), 2630–2643, doi:10.1016/j.atmosenv.2006.11.016, 2007.

Ming, Y. and Russell, L. M.: Predicted hygroscopic growth of sea salt aerosol, J. Geophys. Res. Atmospheres, 106(D22), 28259–28274, 2001.

Murray, C. J., Ezzati, M., Lopez, A. D., Rodgers, A. and Vander Hoorn, S.: Comparative quantification of health risks: conceptual framework and methodological issues, Popul. Health Metr., 1(1), 1, 2003.

Pope, C. A., III, Burnett, R. T., Thun, M. J., Calle, E. E., Krewski, D., Ito, K. and Thurston, G. D.: Lung Cancer, Cardiopulmonary Mortality, and Long-term Exposure to Fine Particulate Air Pollution, JAMA, 287(9), 1132–1141, doi:10.1001/jama.287.9.1132, 2002.

Prather, M., Ehhalt, D., Dentener, F., Derwent, R., Dlugokencky, E., Holland, E., Isaksen, I., Katima, J., Kirchhoff, V., Matson, P., Midgley, P., Wang, M., Berntsen, T., Bey, I., Brasseur, G., Buja, L., Pitari, G. and Et, A.: Chapter 4: Atmospheric Chemistry and Greenhouse Gases, Cambridge University Press. [online] Available from: https://ricerca.univaq.it/handle/11697/24359#.WItxbn0XGao (Accessed 27 January 2017), 2001.

Punger, E. M. and West, J. J.: The effect of grid resolution on estimates of the burden of ozone and fine particulate matter on premature mortality in the United States, Air Qual. Atmosphere Health, 6(3), doi:10.1007/s11869-013-0197-8, 2013.

Ramaswamy, V., Boucher, O., Haigh, J., Hauglustaine, D., Haywood, J., Myhre, G., Nakajima, T., Shi, G., Solomon, S., Betts, R. E., Charlson, R., Chuang, C. C., Daniel, J. S., Del Genio, A. D., Feichter, J., Fuglestvedt, J., Forster, P. M., Ghan, S. J., Jones, A., Kiehl, J. T., Koch, D., Land, C., Lean, J., Lohmann, U., Minschwaner, K., Penner, J. E., Roberts, D. L., Rodhe, H., Roelofs, G.-J., Rotstayn, L. D., Schneider, T. L., Schumann, U., Schwartz, S. E., Schwartzkopf, M. D., Shine, K. P., Smith, S. J., Stevenson, D. S., Stordal, F., Tegen, I., van Dorland, R., Zhang, Y., Srinivasan, J. and Joos, F.: Radiative Forcing of Climate Change, Pacific Northwest National Laboratory (PNNL), Richland, WA (US). [online] Available from: https://www.osti.gov/scitech/biblio/899821 (Accessed 27 January 2017), 2001.

Rao, S., Chirkov, V., Dentener, F., Van Dingenen, R., Pachauri, S., Purohit, P., Amann, M., Heyes, C., Kinney, P., Kolp, P., Klimont, Z., Riahi, K. and Schoepp, W.: Environmental Modeling and Methods for Estimation of the Global Health Impacts of Air Pollution, Environ. Model. Assess., 17(6), 613–622, doi:10.1007/s10666-012-9317-3, 2012.

Rao, S., Pachauri, S., Dentener, F., Kinney, P., Klimont, Z., Riahi, K. and Schoepp, W.: Better air for better health: Forging synergies in policies for energy access, climate change and air pollution, Glob. Environ. Change, 23(5), 1122–1130, doi:10.1016/j.gloenvcha.2013.05.003, 2013.

Riahi, K., Dentener, F., Gielen, D., Grubler, A., Jewell, J., Klimont, Z., Krey, V., McCollum, D., Pachauri, S., Rao, S., van Ruijven, B., van Vuuren, D. P. and Wilson, C.: The Global Energy Assessment - Chapter 17 - Energy Pathways for Sustainable Development, in Global Energy Assessment - Toward a Sustainable Future, pp. 1203–1306, Cambridge University Press, Cambridge, UK and New York, NY, USA and the International Institute for Applied Systems Analysis, Laxenburg, Austria. [online] Available from: www.globalenergyassessment.org, 2012.

Silva, R. A., West, J. J., Lamarque, J. F., Shindell, D. T., Collins, W. J., Dalsoren, S., Faluvegi, G., Folberth, G., Horowitz, L. W., Nagashima, T., Naik, V., Rumbold, S. T., Sudo, K., Takemura, T., Bergmann, D., Cameron-Smith, P., Cionni, I., Doherty, R. M., Eyring, V., Josse, B., MacKenzie, I. A., Plummer, D., Righi, M., Stevenson, D. S., Strode, S., Szopa, S. and Zengast, G.: The effect of future ambient air pollution on human premature mortality to 2100 using output from the ACCMIP model ensemble, Atmospheric Chem. Phys., 16(15), 9847–9862, doi:10.5194/acp-16-9847-2016, 2016.

Stevenson, D. S., Young, P. J., Naik, V., Lamarque, J.-F., Shindell, D. T., Voulgarakis, A., Skeie, R. B., Dalsoren, S. B., Myhre, G., Berntsen, T. K., Folberth, G. A., Rumbold, S. T., Collins, W. J., MacKenzie, I. A., Doherty, R. M., Zeng, G., Van, N., Strunk, A., Bergmann, D., Cameron-Smith, P., Plummer, D. A., Strode, S. A., Horowitz, L., Lee, Y. H., Szopa, S., Sudo, K., Nagashima, T., Josse, B., Cionni, I., Righi, M., Eyring, V., Conley, A., Bowman, K. W., Wild, O. and Archibald, A.: Tropospheric ozone changes, radiative forcing and attribution to

emissions in the Atmospheric Chemistry and Climate Model Intercomparison Project (ACCMIP), Atmospheric Chem. Phys., 13(6), 3063–3085, doi:10.5194/acp-13-3063-2013, 2013.

Stjern, C. W., Samset, B. H., Myhre, G., Bian, H., Chin, M., Davila, Y., Dentener, F., Emmons, L., Flemming, J., Haslerud, A. S., Henze, D., Jonson, J. E., Kucsera, T., Lund, M. T., Schulz, M., Sudo, K., Takemura, T. and Tilmes, S.: Global and regional radiative forcing from 20 % reductions in BC, OC and SO4 - An HTAP2 multi-model study, Atmospheric Chem. Phys., 16(21), 13579–13599, doi:10.5194/acp-16-13579-2016, 2016.

Tang, I. N.: Chemical and size effects of hygroscopic aerosols on light scattering coefficients, J. Geophys. Res. Atmospheres, 101(14), 19245–19250, 1996.

Turner, M. C., Jerrett, M., Pope, C. A., Krewski, D., Gapstur, S. M., Diver, W. R., Beckerman, B. S., Marshall, J. D., Su, J., Crouse, D. L. and Burnett, R. T.: Long-Term Ozone Exposure and Mortality in a Large Prospective Study, Am. J. Respir. Crit. Care Med., 193(10), 1134–1142, doi:10.1164/rccm.201508-1633OC, 2016.

Wang, X. and Mauzerall, D. L.: Characterizing distributions of surface ozone and its impact on grain production in China, Japan and South Korea: 1990 and 2020, Atmos. Environ., 38(26), 4383–4402, doi:10.1016/j.atmosenv.2004.03.067, 2004.

West, J. J., Fiore, A. M., Naik, V., Horowitz, L. W., Schwarzkopf, M. D. and Mauzerall, D. L.: Ozone air quality and radiative forcing consequences of changes in ozone precursor emissions, Geophys. Res. Lett., 34(6), doi:10.1029/2006GL029173, 2007.

Wild, O. and Prather, M. J.: Excitation of the primary tropospheric chemical mode in a global three-dimensional model, J. Geophys. Res. Atmospheres, 105(D20), 24647–24660, 2000.

Wild, O. and Prather, M. J.: Global tropospheric ozone modeling: Quantifying errors due to grid resolution, J. Geophys. Res. Atmospheres, 111(D11), doi:10.1029/2005JD006605, 2006.