Author Comment on egusphere-2026-1367

We thank the two anonymous referees for their comments on our manuscript. In this Author Comment we provide our responses to the comments from the two anonymous referees and describe the modifications to the manuscript resulting from these comments.

The comments from the anonymous referees are reproduced here in red text, our responses are given in black text, and text added to the manuscript is shown here in blue.

In addition to changes in the manuscript resulting from the comments by the anonymous referees, we have also made the following changes to the manuscript.

Matthew Kasoar has been removed as a co-author at his own request. He was originally included in the author list due to his involvement with the future fires emission dataset, but during the course of preparing the manuscript this work was carried out primarily by Elisa Bergas-Masso and Douglas S. Hamilton.

We corrected a small mistake in the handover of biomass burning emissions from GFAS4HTAP to the Future Fires emission dataset (Section 5.2). Instead of the handover happening from 2019-2020, it now happens from 2020-2021.

Figure 1 (HTAP3 source regions) has been updated as described in our Author Comment from 23.04.2026.

Comments from Anonymous Referee #1

This paper describes the experimental design of a new round of multi-model HTAP experiments aimed at quantifying the long-range contributions to ozone, PM, and the deposition of nitrogen and sulphur in order to inform the upcoming revision of the Gothenburg Protocol. The paper is well written and the described protocols are provided in sufficient detail to support the participation by various modelling groups. It will be exciting to see the results of this effort in light of new foci relative to previous HTAP assessments (especially the consideration for methane, wildfires, source-receptor relationships under future emissions scenarios, and the requirement to archive hourly surface ozone mixing ratios). I recommend publication after addressing a few comments below:

Line 51: Could you include a citation for this statement / the recent review of the GP?

We have included a citation for the review of the GP as UNECE (2022).

Line 52: Please remove the extra period.

Done.

Line 506: Was 2015 a climatologically average year? Could source-receptor coefficients and impact metrics vary strongly from year-to-year due to inter-annual meteorological variability? I understand how it would not be possible to ask the modelling centres to repeat these experiments for multiple meteorological years, but some more justification or discussion of these caveats should be included.

We thank the reviewer for pointing out this important aspect of the experimental design which was not addressed thoroughly enough in the original manuscript. The years 2015 and 2040 were chosen as the present-day reference year and future target year respectively for compatibility with the policy process for the revision of the Gothenburg Protocol. At the time of writing, it is expected that emission reduction commitments under the revised Gothenburg protocol will be specified relative to 2015 and should be achieved by 2040. We believe that the choice of 2015 as the present-day reference year should already be mentioned in Section 4.1, where we give the rationale for the choice of 2040 as the base year for perturbation runs. We have added the following text to the first paragraph of Section 4.1.

> The year 2015 is chosen as the historical reference year, as this is the year which is currently being considered as the base year for emission reduction commitments under the revised Gothenburg Protocol.

With regards to the reviewer's question on the climatological representativeness of 2015, we believe that Section 4.1 is also the best place to add text, as this is the section in which the concept of the meteorological reference year is introduced. As the reviewer points out, it is difficult to ask modelling centres for simulations covering multiple meteorological years. This is due to the extra computational requirements of conducting such simulations. As requested, we have added a discussion of the overall representativeness of 2015 as a meteorological reference year to Section 4.1.

> At the time, 2015 was the warmest year on record and was characterised by a strong El Niño event (WMO, 2016). Such large-scale modes of climate variability influence atmospheric circulation and can affect the transport and chemical production of ozone, implying that source-receptor relationships and impact metrics will exhibit some degree of interannual variability (eg. Zhang et al., 2015). Consequently, the source-receptor coefficients presented here should be interpreted as being representative of meteorological conditions in 2015 rather than as invariant quantities.
>
> The use of a single meteorological year represents a compromise between policy relevance, scientific completeness and computational feasibility. The HTAP3-OPNS experiment matrix comprises a large number of perturbation simulations (Section 7.1), making it impractical to repeat every experiment for multiple meteorological years across the participating modelling centres. While the magnitude of individual source-receptor coefficients is therefore expected to depend to some extent on the chosen meteorological year, previous HTAP phases and related source-attribution studies have demonstrated that a single meteorological year nevertheless provides valuable information on the relative importance of emission sources and transport pathways (eg. HTAP, 2010).

Line 515: It could be nice to include a short description of the major regional differences in anthropogenic and biomass burning emissions between GAINS LRTAP CLE 2040 and GAINS LRTAP CLE 2015. Especially since a major point of this new protocol is the calculation of S-R relationships under future emissions scenarios, how much do we expect these relationships to differ under CLE 2040 relative to CLE 2015?

In response to this comment and to a similar comment by the other anonymous referee, we have included a figure showing the evolution of the emissions under the GAINS LRTAP CLE and MTFR scenarios as Figure 1 and Figure 2 respectively.

Line 696: Is it possible that the signal-to-noise ratio under SSP2-4.5 will be too small to quantify a climate penalty by 2050 relative to 2015? Like above, the authors could consider including a brief description of the CO2, CH4, and temperature differences expected under SSP2-4.5 by 2050 relative to 2015.

The use of decadal means, rather than individual years, reduces the influence of interannual meteorological variability and improves the signal-to-noise ratio when quantifying changes driven by emissions and climate. Under SSP2-4.5, the climate signal by mid-century may remain small relative to interannual variability for some regions and metrics. Quantifying the extent to which the climate-driven response under the GAINS LRTAP scenarios emerges above this variability is therefore one of the objectives of these transient simulations.

We have included the following new paragraph in Section 4.2, the overview section for the transient future scenario experiments.

> Under SSP2-4.5, the climate signal by mid-century may remain small relative to interannual variability for some regions and metrics. Quantifying the extent to which the climate-driven response under the GAINS LRTAP scenarios emerges above this variability is therefore one of the objectives of these transient simulations. In 2015 (2050), SSP2-4.5 is associated with atmospheric CO_2 concentrations of approximately 400 (507) ppm (Meinshausen et al., 2020) and CH_4 concentrations of approximately 1834 (2211) ppb (von Caspel et al., 2025). By the 2040–2049 period, SSP2-4.5 is expected to be associated with approximately 0.8°C additional global warming relative to 2015 (IPCC, 2021).

Comments from Anonymous Referee #2

This manuscript describes the requested model configurations, simulations, and output for the HTAP3-OPNS project and is thus an important contribution worthy of publication. These simulations support the upcoming assessment that will guide revisions of the Gothenburg Protocol under CLRTAP with respect to ozone, particulate matter and deposition of nitrogen and sulphur. The documentation of the modeling protocol is critical, as is clearly laying out the questions that the simulations are designed to answer.

We thank the reviewer for their constructive overall assessment of the manuscript.

The current version is written more like a technical report than a typical scientific article. While this is reasonable given the need to document the detailed simulations, some editing would help the reader follow how the simulations will deliver science needed to inform the GP. Merging sections avoids both redundancy (handling of agricultural waste burning, initialization) and very short sections.

We agree that experiment description papers differ in style from conventional scientific articles, as their primary purpose is to provide sufficient technical detail for experiments to be reproduced consistently across modelling groups. Nevertheless, we appreciate the reviewer's suggestion and have revised the manuscript to improve the overall flow and to more clearly explain how the requested simulations address the scientific questions relevant to the revision of the Gothenburg Protocol. The reviewer makes several more concrete suggestions regarding the structure and flow of the manuscript below, where we also give our responses and describe our modifications.

Regarding the discussion of agricultural waste burning, we do not find the existing text redundant. Section 5 describes the input datasets, and we believe it is important in this section to give clear explanations of which datasets are to be used in which experiments. Likewise, Section 7 describes model configuration, and the exact combination of emission datasets is an important aspect of this. Including discussion of agricultural waste burning in both sections decreases the danger that incorrect combinations of emission datasets might be used by some modelling centres.

On the issue of model initialization and other technical details regarding the setup of the model runs, we have removed these details from Section 4 (high level overview of the runs) and moved these into Section 7 (detailed discussion of model setup).

It is suggested to encourage modeling groups to document many choices left at their discretion (for example lines 227-228, 243, 267-269, 301-303). Collecting this information in a publicly available table or website would facilitate future use and interpretation of the simulations. Suggest adding a request to describe nitrate aerosol and SOA in the models. A new section that includes all requested additional documentation may lower the barrier for modeling groups to deliver it. For example, the request for the machine-readable model chemical mechanism is easy to miss (459-461). This section could follow lines 477-480.

We thank the reviewer for this suggestion and agree that this will make life easier for modelling centres as well as subsequent users of the datasets produced during the exercise. We have collected this information and summarised it in a new section as the referee suggests. This new Section 7.5 is placed at the end of Section 7 and a forward reference to this new section is placed at the beginning of Section 7.

The emphasis of HTAP-O3PNS is heavily slanted towards ozone (Table 2). Seeing as SO_2 and NH_3 perturbations are not prioritized, how useful are these runs for long-range transport of PM and deposition of S or N?

Previous HTAP assessments (eg. HTAP, 2010) have found that long-range transport over intercontinental distances is of more importance for ozone than for other, shorter-lived species such as PM and for the deposition of N and S. For this reason, we expect the source-receptor analysis to have its strongest focus on ozone. Given that PM, and the deposition of N and S are also handled by the Gothenburg Protocol, we have also included these in the design (see the "ALL" experiments).

Section 3

1. Given that the emission reference is only to the data files, the impact of this paper would increase by adding figures showing the emissions under these scenarios; even more useful would be to compare them with the SSP2-4.5 so a reader can see they are "broadly consistent".

With "broadly consistent" we are referring to the energy scenario used in the GAINS model during the production of the scenarios as well as the associated CO_2 emissions. In this sense the underlying climate pathway is consistent with SSP2-4.5. We make no claims about the consistency of the air pollutant emissions with those from SSP2-4.5 as the GAINS model contains different representations of emission control technology than the integrated assessment models used in construction of the SSP scenarios. We have modified the text to clarify that it is the underlying climate pathway that is broadly consistent between the GAINS emissions and SSP2-4.5.

In response to this comment and to a similar comment by the other anonymous referee, we have included a figure showing the evolution of the emissions under the GAINS LRTAP CLE and MTFR scenarios as Figure 1 and Figure 2 respectively.

2. HILO: Since a similar sounding scenario is being used in CMIP7 to indicate high-to-low CO_2 emissions (Van Vuuren et al., 2026), the authors might consider renaming to clarify that this is high methane – low air pollution.

The name HILO is already established in this exercise for input datasets and some groups have already submitted runs using this name. At this stage we believe it would add confusion to rename this scenario. We do note the referees concerns though, and have added text at key places in the manuscript to remind readers that this scenario is characterised by high methane and low air pollution.

Section 4. Seeing as modeling groups and future analysts of the data are most likely to want the details of individual sets of simulations, I suggest removing the overview and instead breaking it into substantive individual sections that include the details (combine 4.2 with 7.4, 4.3 with 7.5, etc) to avoid referring to multiple sections with redundant text.

We agree that there is some overlap between Sections 4 and 7. However, the two sections serve different purposes. Section 4 provides a high-level overview of the different experiment families and their scientific objectives, allowing Sections 5 (input data) and 6 (requested output) to refer to the experiments before the detailed protocols are introduced. Section 7 then provides the technical specifications required to perform each experiment.

In response to the reviewer's comment, we have streamlined Section 4 by removing technical details that are repeated later and leaving only the scientific motivation and high-level overview of each experiment. This preserves the overall organisation while reducing redundancy.

4.1 Map experiments to Questions in Section 2 and/or emissions in Section 3?

We have added the following text to Section 4.1.

> The source/receptor relationships derived from the perturbation runs will address the question posed in Section 2 on the intra- and extra-regional contributions to air quality. The suitability of models for quantifying these contributions will be assessed by both detailed evaluation of models against observations (Section 8.2.1) and by the uniformity of the model responses to the perturbations themselves. The diagnostic data requested in Section 6 may help to explain inter-model differences. The construction of an ensemble emulator (Section 8.2.2) will enable the rapid assessment of alternative future scenarios based on the modelled source/receptor relationships.

4.2 What specifically is being assessed about the GAINS LRTAP scenarios? How is climate change represented in the scenario design?

We agree that the text of the first bullet point was not clear, so we have changed this to "Provide a direct assessment of the evolution of air quality under the GAINS LRTAP scenarios.".

Regarding the representation of climate change in the experiment design, we have added the following text.

> Climate change is represented in these experiments through the use of forcing datasets (SSTs, long-lived greenhouse gases, etc.) corresponding to the SSP2-4.5 scenario from CMIP6.

4.2 and 7.4.1 The questions driving the recommendations for the Atmosphere-Only runs in Table 4 need to be stated more clearly. Those questions should determine the required specifications. It is not yet clear whether the design will accomplish the intended outcomes. For example, the requested decoupling of composition is confusing given the recommendation that modelers use decadally averaged SSTs and sea ice as in a CMIP6-AMIP configuration. Since the SSTs can't respond to the radiation scheme in this setup and the SSTs are the dominant driver on the atmospheric climate, why the need to decouple from radiation? If the goal is to force the model to the same meteorology, why not nudge?

We thank the reviewer for this comment and agree that the scientific questions motivating the Atmosphere-Only simulations need to be stated more clearly. The purpose of the clim-2015 configuration is not to constrain the model to reproduce a particular sequence of meteorological conditions, but rather to maintain a stationary climate representative of approximately the period 2010-2020 while allowing the atmosphere to generate its own internal meteorological variability. Prescribing climatological SSTs and sea ice representative of this period provides the principal constraint on the large-scale climate state while avoiding a long-term climate-change signal.

Although prescribed SSTs prevent the slower SST-mediated climate response, changes in atmospheric composition can still produce rapid atmospheric and land-surface adjustments, including changes in clouds and circulation. Decoupling composition from radiation is therefore intended to suppress these fast responses as well as the slower climate response.

For this reason, we do not propose nudging the atmosphere towards specified meteorological fields. Such nudging would address a different objective, namely ensuring that simulations experience approximately the same meteorological sequence, and would constrain atmospheric variability which we wish the models to represent freely.

We have added the following text to the high-level overview of the future transient simulations in Section 4.2 as justification for the experimental design given in Section 7.3.1 (previously Section 7.4.1).

> In order to separate the effects of climate change from the effects of changing emissions on future air quality, we perform an additional set of runs using the GAINS LRTAP scenarios for the period 2010-2050, but with a stationary climate representative of approximately the period 2010-2020, allowing models to generate their own internal meteorological variability. To ensure that the GAINS LRTAP scenarios themselves do not lead to fast responses in modelled climate, this additional set of runs will be performed with composition decoupled from radiation.

4.3 What historical global emission inventory is used? Elaborating on the evaluation plan in 8.2.1 including some discussion of datasets would be helpful. Will COVID period emissions and responses be used as a test of the models?

In Section 4.3 we have now explicitly mentioned that the HTAPv3.2 inventory will be used.

In Section 8.2.1 we have added the following sentence.

> The harmonised datasets collected and made available by the TOAR-II (Tropospheric Ozone Assessment Report phase 2, https://toar-data.org/) will be especially useful for the evaluation of ground-level and tropospheric ozone.

Regarding the COVID period, this will only be of limited usefulness, since many of the simulations in HTAP3-OPNS do not cover this period. The perturbation runs (Section 7.1) use 2015 as a present-day base year, and the GAINS LRTAP scenarios do not include the effects of the COVID lockdowns.

Section 5. Merge with Section 3 so that a reader can easily find all emissions information?

We believe that Sections 3 and 5 serve different purposes. Section 3 introduce the GAINS LRTAP scenarios in their context as the foundation for scientific assessments supporting the revision of the Gothenburg Protocol. The early introduction to these scenarios avoids the need for forward references in the high-level overview descriptions of the model experiments in Section 4. In Section 5, the full set of input datasets is described in detail. For the GAINS LRTAP scenarios (Section 5.1), this includes details such as the emission sectors present in the dataset and how the deficiencies in the dataset (eg. The lack of aircraft emissions) are handled in general in the HTAP3-OPNS exercise.

Section 6. A table to clarify what is needed for the Edwards and Evans (2017) approach would be helpful either here or in the supplemental spreadsheet. Are any idealized tracers (age of air, biomass burning, etc) requested to aid interpretation of the runs?

The reference to Edwards and Evans (2017) is outdated and should have been removed from the manuscript before submission. While we originally intended to use this approach, subsequent discussions with the first author of this paper as well as the modellers interested in contributing to the HTAP3-OPNS exercise determined that this would be impractical. The approach used by Edwards and Evans (2017) is enabled by specific features of the GEOS-Chem model which are not commonly available in other models.

We have removed the reference to Edwards and Evans (2017). The requested terms for determining chemical budgets are described fully in Section 6.4.

The data request includes a tracer with a 25-day lifetime set to the annual mean of the CO fire emissions in space and magnitude. We expect that this will be primarily of interest for interpretation of the experiments in the HTAP3-Fires exercise.

Section 7:

The organization of this section is confusing. Aren’t the simulations described in section 7.2 a subset of the 14 discussed in 7.1? Why are regional perturbations under global?

We thank the reviewer for pointing this out, and agree that the organization of this section is confusing. We believe that the confusion stems from incorrectly using a level 2 heading (7.2) for the global perturbation experiments. As the reviewer points out, these experiments are indeed a subset of the runs described in Section 7.1. We have therefore changed the heading on Section 7.2 to a level 3 heading (now Section 7.1.2).

The regional perturbation runs described in Section 7.2.1 (now Section 7.1.3 after the change in the organization described above) are performed with global models, so they do belong in the section on global modelling. We have added a short clarification of this to the new Section 7.1.3.

Please provide more background on this HTAP3 ensemble emulator. Is it ozone-only? How will it support the GP revision?

The reviewer makes a similar comment below (referring to line 790 of the submitted manuscript) by making modifications to Section 8.2.2, where the emulator is discussed. In this section we now clarify that the emulator is not ozone-only. We also add the following sentence.

> Use of the emulator will allow rapid assessment of additional mitigation scenarios in addition to the CLE and MTFR scenarios in support of the revision of the Gothenburg Protocol.

We have also added a forward reference to Section 8.2.2 from Section 7.1.

How is NOx from fertilizer handled in the inventories and in the perturbation simulations?

Agricultural activities including fertilizer application are considered in both the GAINS LRTAP (Klimont et al., 2025) and HTAPv3.2 (Guizzardi et al., 2025) inventories. Emissions of NOx and NH_3 are derived from these activities in the GAINS model using standard emission factors, and in the HTAPv3.2 inventory using the respective approaches from the component regional inventories, gap-filled with the global EDGARv8.1 inventory.

In the perturbation runs, perturbations of the NOx from fertilizer application are included in all “NOX” and all “ALL” simulations.

In order to avoid double-counting these emissions, we have added the following text to Section 5.1, and equivalent text to Section 5.4.

> Since the GAINS LRTAP emission datasets include emissions associated with the application of fertilizer, modellers should ensure that their choice of soil NOx emission inventory does not also include this source.

From a policy standpoint, would it be more relevant to increase aviation and biomass burning sources rather than decrease them?

For aviation emissions, the CMIP6 emission dataset used here already includes projected growth in emissions. Negative perturbations in the aviation emissions can be thought of as a way to understand the model responses to mitigation of this emission sector.

Biomass burning emissions on the other hand are kept constant at 2015 levels in all CTM simulations and undergo a negative perturbation in the FIRE simulation. The reviewer has a good point that a positive perturbation may have been more appropriate in this case, given that wildfires and associated emissions are widely expected to increase in the future.

While a negative perturbation of biomass burning emissions will still allow the calculation of the models' sensitivity to this emission sector, we agree that our choice of negative perturbation represents a missed opportunity to assess the model sensitivity in a directionally more likely way. Given that many groups have already started their model runs for this exercise, and that biomass burning is not the focus of this exercise (the FIRE simulation has priority 3 in Table 2), we do not feel that it is necessary to change the design of the experiment in this case.

Any prioritization between the source-receptor versus transient simulations?

We do not specify any prioritization between the different work streams. We expect that modelling centres will choose the set of runs which most closely corresponds to their capabilities, capacity, and scientific interests.

Section 7.3.2. Why only specify methane at lateral boundaries rather than throughout the regional domain? Why the focus on Europe?

The recommendation to specify methane at lateral boundaries is made to ensure consistency between the regional model simulations, so that all regional models have effectively the same source of methane. If methane were to be kept fixed throughout the regional model domains, then differences in OH levels between models could lead to higher effective sources in models with higher OH.

The focus on Europe was determined primarily by the number of expressions of interest in simulating this region in the early planning stages of the exercise, but also because the main focus of the Gothenburg Protocol revision is the EMEP region.

Section 9. It would be useful to give a very brief overview of what was learned from earlier efforts to motivate the need for this new effort. Please provide references to overview papers for these MIPs as the current text is only useful to those who already know them.

In many cases, the other multi-model exercises refer to ongoing or planned work (HTAP3-Fires, MCHgMAP, MethaneMIP). In all of these cases we have already provided references to the relevant overviews and discussed how HTAP3-OPNS is differentiated from or complementary to these exercises.

For the HTAP phase 1 and 2 exercises, we have added references to HTAP (2010) and Turnock et al. (2018). We have also added the information that many modelling centres struggled to provide the full set of requested perturbation runs and that we have attempted to limit the number of requested runs in HTAP3-OPNS.

For AerChemMIP and CCMI we have included references from Zanis et al. (2022) and Abalos et al. (2020) which provide a motivation for using climate models in HTAP3-OPNS by adding the following text to Section 9.4.

> Previous multi-model chemistry–climate studies have demonstrated that climate change can substantially modify future surface ozone independently of changes in anthropogenic precursor emissions. Analysis of AerChemMIP simulations (Zanis et al., 2022) found a climate-driven decrease in surface ozone over much of the globe, primarily associated with enhanced ozone destruction under increased atmospheric water vapour, but a climate penalty in polluted regions and regions influenced by enhanced natural VOC emissions, where warming increased the efficiency of ozone production from precursor emissions. CCMI simulations have additionally shown that climate change is expected to enhance stratosphere-to-troposphere ozone transport, with the magnitude of this effect increasing under stronger climate forcing (Abalos et al., 2020). These results demonstrate that the response of surface ozone to future emission controls depends on the evolving climate and motivate the explicit consideration of climate change in the HTAP3-OPNS future scenario experiments.

Section 10. Something like the last paragraph would be helpful context earlier in the paper.

We thank the reviewer for this suggestion, and agree that this would help to introduce the reader to the HTAP3-OPNS exercise. We have added the following text to the last paragraph of the introduction.

> The experiments have complementary roles within the overall HTAP3-OPNS framework. Perturbation simulations using Chemical Transport Models and pre-existing emission scenarios will be used to quantify source–receptor relationships and develop an ensemble emulator for exploring mitigation options, while transient Chemistry-Climate Model simulations will investigate the influence of climate change on future air quality under policy-relevant air pollution control scenarios. Simulations of the recent historical period using a state-of-the-art global mosaic emission inventory will support model evaluation and provide connections with concurrent activities addressing total atmospheric deposition and the impacts of wildfires on air quality.

Detailed comments.

Lines 149-151. State that tagging and adjoint approaches are *in addition* to the standard perturbation runs to allow apples-to-apples comparison?

We have clarified that the use of these alternative source attribution techniques is strictly in addition to the standard runs.

Lines 285-287. Include emission and concentration time series in a figure?

We have included a figure of the methane concentrations from the CLE, MTFR, and HILO scenarios in this section.

Lines 292-293. Why 5 years of spinup if methane and hydrogen concentrations are fixed?

The text specified "up to" 5 years of spinup. We agree that this is a rather large amount of spinup for CTM experiments with methane and hydrogen concentrations fixed. Modelling centres are expected to know more about how much spinup time their models need. We have given some guidance on this in Section 7.1 for the perturbation experiments. While responding to this comment we also realised that we could explicitly refer to spinup time in the first paragraph of Section 7 as a model configuration parameter that we leave to the expert judgement of each modelling centre.

Line 401. Time scale of this request for total atmospheric deposition?

We have clarified in the text that output is requested at monthly mean resolution.

Lines 472-474. Can ERF be calculated from only knowing SARF and IRF? I thought the CMIP6 ERF protocol relied on targeted simulations with fixed SSTs as described in Pincus, Forster & Stevens (2016).

We thank the reviewer for pointing this out. We agree that the wording in the manuscript was misleading. The proposed diagnostic double radiation calls allow the instantaneous radiative forcing associated with aerosols or ozone to be diagnosed, and stratospherically adjusted radiative forcing may additionally be provided by models capable of diagnosing it. However, effective radiative forcing cannot in general be calculated from IRF and SARF alone, since ERF also includes rapid tropospheric and land-surface adjustments. In CMIP6/RFMIP and AerChemMIP, ERF was instead diagnosed using dedicated atmosphere-only simulations with fixed SSTs and sea ice, allowing rapid adjustments while suppressing the slower SST-mediated climate response (Pincus et al., 2016; Collins et al., 2017). We have therefore deleted the final sentence of Section 6.6 to remove the implication that the proposed diagnostic output is sufficient to calculate ERF.

Line 625. Consider adding a separate request in the supplemental spreadsheet for this high frequency output

Given that the boundary condition fields supplied by the global models to the regional models will be quite large and can therefore not be hosted on the Aerocom server, and that only a very small subset of the global modellers are expected to produce and host these, the decision was made not to formally specify these fields as part of the data request.

We have added a short clarification to this section.

Line 751. Provide a specific lat/lon resolution recommendation?

There is no requirement on a specific model resolution for any of the model simulations in HTAP3-OPNS. This is mentioned in the first paragraph of Section 7.

In response to this comment, we have added some text here stating that the regridding should be done at an equivalent resolution.

Line 790. Is this only ozone or also S, N deposition and PM?

While the main focus of the emulator will be ozone (see our earlier response to the comment on the main emphasis of the HTAP3-OPNS exercise), the emulator will also be constructed for PM and the deposition of N and S. We have modified this paragraph in Section 8.2.2 to explicitly include these other species, while acknowledging that previous HTAP work on this has focused on ozone.

Lines 803-805. How will the reliability of these results be assessed and communicated?

The reliability of these results will be assessed through the evaluation of model results for the present-day period as described already in Section 8.2.1. All work streams include simulations covering either the year 2015 or the period 2010-2020 which can be compared against observations to assess the model performance. Model evaluation will be the initial focus of the analysis in HTAP3-OPNS and a standalone publication is planned with the results of this evaluation.

Lines 831-833. Are emitted VOC like acetone that photolyze to produce OH included?

Yes. As mentioned in the data request, all reactions producing OH should be included in the calculation of the OH production term.

Lines 862-863. Hourly surface fields were requested in earlier assessments (e.g., Reidmiller et al., 2009); the “aerhourly” table in the supplemental file does not include hourly deposition fluxes for N and S species?

The omission of hourly N and S deposition fluxes from the aerhourly table is intentional. While hourly surface concentrations are useful for evaluating pollutants with strong diurnal variability and for calculating relevant air-quality metrics, the analysis of N and S deposition in HTAP3-OPNS focuses on accumulated deposition over longer averaging periods. We therefore consider the monthly deposition fluxes sufficient for the planned deposition evaluation and source–receptor analysis. Requesting the corresponding deposition fields at hourly resolution would substantially increase the output volume without providing a clear benefit for the planned analyses.

We have modified the text in Section 9.1 to clarify that hourly deposition fluxes are only requested for ozone.

Table 1. What is Met.no? The Hamilton et al. reference is missing.

We have completely updated Table 1 to be a native Word table instead of an inserted graphic in anticipation of this being necessary during typesetting. The original content is preserved. The Hamilton et al. reference is updated to Bergas-Masso and Hamilton (2026), and the table caption has been expanded to include references to the sections in which each dataset is described, including the Met.no methane concentration data.

Comments on requested diagnostics listed in the supplemental file:

We thank the reviewer for their detailed comments on the data request. Given that many modelling centres have already begun working on their simulations for this exercise, we do not feel that it is appropriate at this stage to make major modifications to the data request. This goes especially for including any new fields in the data request. In response to the reviewers comments we focus on explaining our choices. Modifications to the data request itself are restricted to the removal of fields which are clearly not needed and which the models contributing to HTAP3-OPNS will not be providing in any case.

The file provides important information not only for the current modellers but also for those using the modelled datasets in the future. With this future use in mind, it seems important to document the scientific rationale for the aerzonal, aerzonal-vert, ModelLevelAtStations and “List of TOAR2 HEGIFTOM” stations worksheets in the manuscript.

As described in Section 6, the data request in HTAP3-OPNS is based on the data request from the earlier AerChemMIP exercise. The AerChemMIP data request itself borrowed from previous model intercomparison exercises, including HTAP2. We believe that keeping the majority of the worksheets intact allows us to both maintain continuity with these previous exercises and to facilitate future analyses of the HTAP3-OPNS model archive that may extend beyond the analyses currently planned. We have added the following text to the first paragraph in Section 6 to make this clear.

> The output request retains diagnostics identified as useful in previous model intercomparison and assessment exercises, both to maintain continuity with these exercises and to facilitate future analyses of the HTAP3-OPNS model archive that may extend beyond the analyses currently planned.

The aerzonal and aerzonal-vert sheets fall into this category of inherited sheets. These sheets facilitate comparison of the large-scale meridional and vertical distributions of atmospheric composition across models. For the ModelLevelAtStations and "List of TOAR2 HEGIFTOM stations" we have added these sheets to enable model evaluation. The ModelLevelAtStations sheet comes from HTAP2, and the list of TOAR2 HEGIFTOM stations comes from the more recent TOAR2 assessment. We have added the following text.

> In addition to the worksheets based on AerChemMIP, we have added a request for vertical profile output on model levels at the grid cells corresponding with the locations of ozonesone measurements from the TOAR2 HEGIFTOM dataset (Van Malderen et al., 2025) for the purposes of model evaluation.

aerhourly: Is jNO2 an appropriate proxy for ozone production or would it require additional variables to solve for steady state? Why not request o3prod or tropdo3chm instead?

The reviewer is correct that jNO2 on its own is not enough to completely diagnose ozone production. It is retained in the data request for hourly surface fields because it is usually available from most models and provides useful information about photolysis and shortwave radiation in general. For analysis of tropospheric ozone budgets, we already ask for o3prod, o3loss, do3chm, tropdo3chm and other related fields in the aermonthly-3d sheet.

aermonthly_3d: Artificial tracers (aoanh, cofire25d, and o3ste) are requested but not documented in the manuscript. Are long-lived greenhouse gases emitted? If not, why are 3D distributions being requested?

The named fields are retained in the data request to maintain continuity with previous model intercomparison exercises and to facilitate future analyses of the HTAP3-OPNS model archive that may extend beyond the analyses currently planned. Regarding long-lived greenhouse gases, we expect that some models contributing to the future transient set of runs will use methane emissions (as described already in Section 7.3) so there is a clear scientific rationale for including methane in the requested output. As a tracer with active chemistry in the stratosphere, N2O may also provide useful information about the models. For CO2, retaining this field in the data request serves as a quality control check on how models have kept this tracer fixed.

Is it possible to indicate which variables are most critical for a model to produce? Are some of the requests such as PAH for other HTAP projects? Can the variables for HTAP3-OPNS request be differentiated here?

We already have the following text at the top of Section 6: “Diagnostic fields of particular importance, as well as new fields which have been added for HTAP3-OPNS, are described in this Section below”. The subsections in Section 6 go on to describe the most critical output fields.

The concentrations and deposition fields of various POPs were indeed added to this data request at a time when we were considering a unified data request for HTAP3-OPNS and HTAP3-Fires. We have removed these fields from the data request in response to this comment.

Are monthly deposition fluxes sufficient for meeting the policy needs for N and S deposition?

Acidification and eutrophication result from long-term accumulated deposition to ecosystems. Monthly deposition fluxes are sufficient for HTAP3-OPNS.

If PPFD is not available, would providing total downwelling shortwave radiation be useful?

We request surface downwelling photosynthetic photon flux at hourly resolution from models that can provide this, as this can be used directly in the calculations of the flux-based ozone impact metric PODy. For models that can’t provide this output, we have several options for substituting this data. One option is to use data from another model, as this data is strongly linked to the meteorological conditions, and all models should be using 2015 meteorology. Another option is to estimate this field using other output. For this we could use a combination of the total downwelling shortwave radiation at the surface, which we request at daily frequency, and the jNO2, which we request at hourly frequency.

We have not requested total shortwave radiation at hourly frequency as we want to limit the number of hourly fields, which take up substantial storage space. As mentioned in a previous response, most models should be able to provide the jNO2 output.

References:*Pincus, R., Forster, P. M., and Stevens, B.: The Radiative Forcing Model Intercomparison Project (RFMIP): experimental protocol for CMIP6, Geosci. Model Dev., 9, 3447–3460, https://doi.org/10.5194/gmd-9-3447-2016, 2016.*

Reidmiller, D. R., Fiore, A. M., Jaffe, D. A., Bergmann, D., Cuvelier, C., Dentener, F. J., Duncan, B. N., Folberth, G., Gauss, M., Gong, S., Hess, P., Jonson, J. E., Keating, T., Lupu, A., Marmer, E., Park, R., Schultz, M. G., Shindell, D. T., Szopa, S., Vivanco, M. G., Wild, O., and Zuber, A.: The influence of foreign vs. North American emissions on surface ozone in the US, Atmos. Chem. Phys., 9, 5027–5042, https://doi.org/10.5194/acp-9-5027-2009, 2009.s

Van Vuuren, D. P., O'Neill, B. C., Tebaldi, C., Sanderson, B. M., Chini, L. P., Friedlingstein, P., Hasegawa, T., Riahi, K., Govindasamy, B., Bauer, N., Eyring, V., Fall, C. M. N., Frieler, K., Gidden, M. J., Gohar, L. K., Högner, A., Jones, A. D., Kikstra, J., King, A., Knutti, R., Kriegler, E., Lawrence, P., Lennard, C., Lowe, J., Mathison, C., Mehmood, S., Nicholls, Z., Prado, L. F., Zhang, Q., Rose, S. K., Ruane, A. C., Sandstad, M., Schleussner, C.-F., Seferian, R., Sillmann, J., Smith, C., Sörensson, A. A., Panickal, S., Tachiiri, K., Vaughan, N., Vishwanathan, S. S., Yokohata, T., Zecchetto, M., and Ziehn, T.: The Scenario Model

Intercomparison Project for CMIP7 (ScenarioMIP-CMIP7), Geosci. Model Dev., 19, 2627–2656, https://doi.org/10.5194/gmd-19-2627-2026, 2026.