Answers to reviews of GMD-2021-64 “Description and evaluation of the tropospheric aerosol scheme in the Integrated Forecasting System (IFS-AER, cycle 47R1) of ECMWF”

The authors would like to thank the three reviewers for their thorough review and constructive comments. They helped a lot in improving the manuscript. The article has been modified in depth to try to address the comments from all three reviewers. In particular, the structure of the paper has been reviewed so as to better separate the model description and evaluation parts. We added a subsection to describe the observational datasets used for evaluation purposes. A new subsection to summarize skill scores for the key parameters (AOD, PM) for the three experiments has also been added. A large fraction of the plots have also been redrawn.

The reviewer comments are shown below in *blue italic*; our answers are shown in black.

Reviewer #1

This manuscript presents a description of the changes carried out between C47R1 and C57R1 in the aerosol module of IFS. My main concern is about writing quality. In its current state, the manuscript is very difficult to read and not very well structured. Many information is not provided in the manuscript as often only the changes made in the model are described with differences with parameterization used operationally and not making the paper confusing. For examples, almost no indications are given on how the formation of secondary aerosol was taken into account in the model. The hypotheses of the model are often not clearly explained. The manuscript is more a technical note designed for people working with the IFS model and not a scientific paper for the whole scientific community. In my opinion, the manuscript should not be accepted in its current state. I strongly advise the authors to make significant revisions of the manuscript, to present well the parameterizations by explaining their physicochemical significance and the assumptions behind these parameterizations. The authors should write a standalone manuscript with all the information necessary to understand the model.

We agree that the structure and layout of the manuscript need to be improved. The whole manuscript has been restructured so as to better separate the description and evaluation parts. Regarding the description of the formation of secondary aerosols, more detail has been added for the sulfate formation: the following paragraph has been added:

> “The Sulphur chemistry in CB05 is as described in Huijnen et al., (2010). In short, in total 111 Tg SO2 is emitted, which is composed of 97 Tg anthropogenic, 13 Tg volcanic and 1 Tg biomass burning emissions. In addition 38 Tg dimethyl sulfide emissions are applied taken from climatological values, which is oxidized to form SO_2 (37 Tg) and the rest (5.3 Tg) methyl sulfonic acid.

> This leads to an annual production of SO_4 of 124 Tg, both through gas-phase oxidation with OH and aqueous-phase oxidation including reactions with H_2O_2 and O_3."

For nitrate and ammonium production, this was described in detail in our 2019 GMD paper, and no change have been brought in cycle 47R1. The following sentences have been added to summarize the main formation processes (gas/particle partitioning and heterogeneous reactions):

The gas-particle partitioning scheme estimates nitrate and ammonium production through the neutralization of HNO3 by theNH3 remaining after neutralisation by sulfuric acid:

NH3 + HNO3 ↔ NH4NO3

The formation of nitrate from heterogeneous reactions of HNO3 with calcite (a component of dust aerosol) and sea-salt particles is accounted through the following reactions:

HNO3 + NaCl –→ NaNO3 + HCl

2HNO3 + CaCO3 –→ Ca(NO3)2 + H2CO3

No explanation is given in the choice of the parameters in Table 2. What is the basis for these parameters? I think there are wrong values given. For some lines, several r_{mod} *are given with no explanation or on the opposite not enough values for DUST or salt lines where there is several modes of aerosols. Are the parameters corresponding to a volume distribution or a number distribution? The values of* r_{mod} *seem to be very low (often around 10 nm, as if the aerosol would all be nanoparticles in the atmosphere). It does not seem realistic as it is not in the accumulation mode as it should be. In that case, all the processes calculated with the aerosol diameter are probably not well represented in the model. However, later in the text, it is explained that the deposition rates are not computed with the aerosol size. The density of organic aerosol seems very high.*

These parameters are mostly similar to those presented in Reddy et al. (2005). We agree that some values are not in line with size distribution estimates from more recent work; revisiting them is the subject of ongoing work. New values and associated optical properties will hopefully be used for cycle 49R1. However, the values used presently are still of interest. We agree that this table needs more explanation, and thank the reviewer for this suggestion. Only sea-salt aerosol and coarse mode nitrate use two modes for the underlying size distribution assumption, but sea-salt aerosol and dust each have three bins. What is shown is a **number distribution**, not a volume distribution; number distribution is more commonly used as an input to codes that compute aerosol optical properties using Mie theory.

The sentence

“The aerosol species and the assumed size distribution are shown in Table 1”

Has been changed into:

“The aerosol species and the assumed number size distribution are shown in Table 1”

And in the table legend, the following sentence has been added:

“The number size distribution is assumed to be monomodal for all species except sea-salt and coarse mode nitrate for which a bimodal size distribution is assumed.”

Also, in the conclusion, the following sentence has been added:

“Also, some size distribution assumptions are outdated and need revision”

Almost no precision is given in the representation of phenomena. Many times, the authors refer to the previous article, without giving a short description of what the model does and what are the assumption. I think the paper should be entirely restructured as it is often not clear what is done and what is the rationale behind the parameterizations. Here are a few examples :

- *The description of deposition rate arrives late in the text (section 5) whereas global budget including deposition are given quite early. It would help to provide all the information of the model earlier.*

  Thank you for the suggestion; the structure of the manuscript has been modified so as to avoid this kind of problem.

- *How was the partitioning of nitric acid and ammonia represented? Has a thermodynamic module been used? How is the formation of coarse nitrate represented in the model?*

  As mentioned earlier, a summary of how these processes are represented has been integrated into the manuscript. Additionally, the first sentence of the section has been rephrased so as to indicate that we used the simplified thermodynamical module EQSAM:

  “The production scheme of nitrate and ammonium through gas/particle partitioning processes, and of nitrate from heterogeneousreactions on dust and sea-salt particles, is detailed in Rémy et al. (2019) and has been adapted from Hauglustaine et al. (2014),
  which uses the equilibrium simplified aerosol model (EQSAM, Metzger et al. (2002)) approach. »

- *Section 2.2. What are the heterogeneous reactions included in the model?*

They consist in the heterogeneous reactions of nitric acid on calcite and on sea-salt aerosol particles. The following sentences have been added:

"The formation of nitrate from heterogeneous reactions of HNO3 with calcite (a component of dust aerosol) and sea-salt particles is accounted through the following reactions:

HNO3 + NaCl –→ NaNO3 + HCl

2HNO3 + CaCO3 –→ Ca(NO3)2 + H2CO3"

- *SOA emissions scaled to CO? While I personally think that such approaches are not reliable, I understand that some models use this kind of approach for simplification purposes. It should be explained how these emissions are determined and what kind of chemistry it implicitly accounts for. What are the consequences of using this kind of simplified parameterizations? Are there SOA/CO ratio specific of emissions sectors (I don't understand how a single ratio could be used)?*

We agree that this is a very much simplified approach. This is meant to represent the anthropogenic SOA component: biogenic SOA uses direct emissions that are scaled on natural terpene emissions. A more complete discussion of this choice is included in the 2019 paper and has not been repeated here, as this part has not changed between cycle 45R1 and 47R1. In their 2011 paper, Spracklen et al. found that they achieved best results in simulating secondary organic aerosols when they assumed a large SOA source (100 Tg per year) from sources that matched anthropogenic pollution, and more specifically CO, although they couldn't explain this by chemical formation pathways. Using this approach helps a lot into improving the simulated AOD and PM over polluted areas. In the "operational configuration section", the following sentence has been modified to add more detail:

"Since cycle 43R1, as detailed in Remy et al. 2019, direct anthropogenic SOA emissions scaled on anthropogenic CO emissions are added to organic matter emissions. This large anthropogenic SOA source derives from the work of Spracklen et al. 2011 who found that they achieved best results in simulating secondary organic aerosols when they assumed a large SOA source (100 Tg per year) from sources that matched anthropogenic pollution. Biogenic SOA emissions that are taken as a 15% fraction of natural terpene emissions following \citet{dentener:06} are also added to organic matter emissions"

In cycle 47R1 IFS-AER, emissions are used in bulk (sum of all sectors): no sectoral information are used. A single scaling factor of 0.15 was used.

- *Section 5.2.1. I don't understand how the Di parameter was selected. Should the evaporation of droplets lead to an "evaporation" (probably not the good term as I don't see how non-volatile dusts or BC can evaporate) of particles from the droplet? In reality, it would probably stay in the droplet unless the evaporation of the droplet is complete or has a low settling velocity.*

  The Di parameter represents the in-cloud scavenging coefficients, defined as the fraction of the aerosol in the cloud part of the grid box that is embedded in the cloud liquid/ice water. Evaporation doesn't impact this parameter (but it is represented in the re-evaporation process that is detailed a bit below in the same section). The values have been derived from Reddy et al. (2005) and Stier et al. (2005). The following sentence has been modified to add more explanation:

  "where Di are the in-cloud scavenging coefficients, defined as the fraction of the aerosol in the cloudy part of the grid boxthe aerosol in the cloudy part of the grid box that is embedded in the cloud liquid/ice water."

  The following sentence has been modified:

  "for water droplets, the values of D_w from Table 9 are used, which have been derived from Reddy et al. (2005) and Stier et al. (2005)

- *I don't understand the basis behind the equation in section 5.2.2 with some not defined parameters.*

  Apologies for the non defined parameter f_k, for which a definition has been added. This subsection has been entirely rewritten to provide a summary and a better explanation of the Luo et al. (2019) approach for in-cloud scavenging.

- *Section 6.2: Before comparing to AeroCom, it would be useful to provide information on the exercise. It is very difficult to understand what is done without having basic information on the exercise. What are the conditions of the simulations? When comparing the flux of emissions for Sea Salts, are the emissions corrected by the factor 4.3? Are all the emissions computed for the same conditions of humidity or are they corrected the same way? I understand the idea of the AeroCom exercise to compare the budgets between the model to see how much results differ between models. I however fail to see the interest of evaluating the representation of the budget from IFS-AER by comparison to the median of models. All models could fail to represent one process, in that case the median of models would be wrong.*

  Indeed, comparing to AeroCom, for a different year (2010 vs 2017) and for simulations using different emission datasets is not a validation, it just gives an idea on how IFS-AER compared against other global aerosol models. The fast that IFS-AER also participated to AeroCom phase III provides this kind of feedback: the interested reader may refer to Gliss et al. (2021) for more information. A paragraph has been added to introduce the AeroCom initiative, and also to state that this comparison is not a validation :

“The values indicated in Table 11 can be compared against the values from the AeroCom Phase III control experiment, as reported in Gliss et al. 2021, which also includes data from IFS-AER cycle 46R1. The objective of the AeroCom initiative is to document differences of aerosol component modules of global models and to assemble data-sets for model evaluations. 14 global models participated to the Phase III control experiment, which consisted in simulating aerosols for the years 1850 and 2010. All models used the same CMIP6 emissions. Comparing against the Aerocom median doesn't validate or invalidate IFS-AER results; rather it gives an idea on how IFS-AER stands as compared to other global aerosol models. The AeroCom median below refers to the 2010 experiment.”

- *Comparison to observations: why did you keep traffic stations in the analysis. At the resolution of the simulations, it induces a very large bias. All the comparison should be redone without including traffic stations (or removed from the paper).*

Thanks for the suggestion. In Figure 17, the panels for North America have been replaced so that the plots include only background rural stations. For Chinese air quality, a selection has likewise been done so that only background rural stations are taken into account.

- The aqueous-phase sulfur chemistry seems to be lacking from the model (as only the gas-phase chemistry of CB05 is mentioned) while it is generally the main oxidation process of SO2 and of sulfate formation. I don’t see how the sulfate aerosol formation could seem trustworthy. The fact that IFS-AER overestimates sulfate concentrations without accounting for the aqueous-phase chemistry seem to indication that the representation of some phenomena are not accurate.

IFS-CB05 does include aqueous chemistry, and formation of sulfate through aqueous chemistry reactions is represented. This has been added in the section that describes the sulfur cycle:

“The Sulphur chemistry in CB05 is as described in Huijnen et al., (2010). In short, in total 111 Tg SO2 is emitted, which is composed of 97 Tg anthropogenic, 13 Tg volcanic and 1 Tg biomass burning emissions. In addition 38 Tg dimethyl sulfide emissions are applied taken from climatological values, which is oxidized to form SO_2 (37 Tg) and the rest (5.3 Tg) methyl sulfonic acid.

This leads to an annual production of SO_4 of 124 Tg, both through gas-phase oxidation with OH and aqueous-phase oxidation including reactions with H_2O_2 and O_3.”

The reviewer is probably referring to Figure 20, which compares the observed (CASTNET) and simulated (IFS-AER CY45R1 and CY47R1) sulfate surface concentration. Overall, for all stations and the year 2017, there is no sign of a systematic positive bias for the CY47R1_NEWDEP simulation. However, there is a strong positive bias for the CY45R1 simulation, which is not coupled to IFS-CB05: for this simulation, the sulfate production is computed by a very much simplified method, using latitude as a proxy for the abundance of OH radical and fixed lifetimes at the Poles and the Equator.

Minor comments:

"The mass mixing ratios of these two are passed from IFS-CB05 to IFS-AER, used in the nitrate and ammonium production schemes, and updated in return by those schemes." Are IFS-CB05 and IFS-AER, two separate models. In that case, it would be necessary to have a scheme. Otherwise, I suggest to change or remove this, sentence as I think the authors just want to say that the concentrations given the gas-phase chemical mechanism are used as inputs for the aerosol module.

AER and CB05 correspond to the tropospheric aerosol and chemistry schemes that have been integrated into the IFS, so the terms "IFS-AER" and "IFS-CB05" denote two different components of the same model, which interact through mass mixing ratios of various species (SO_2, SO_4, HNO_3, NO_3, NH_3, NH_4 mainly in cycle 47E1). The following sentence has been added in section 1 to clarify this:

"IFS-AER is run by default coupled with the operational Carbon Bond 2005 (CB05, Yarwood et al. 2005} tropospheric chemistry scheme that has been integrated into the IFS (Flemming et al. 2015) and is afterwards denoted "IFS-CB05".

About this particular sentence, the reviewer is correct: the concentrations of the gas-phase nitric acid and ammonia are provided the the tropospheric chemistry scheme and used as input of the aerosol module. Following the gas/particle partitioning and heterogeneous reactions, the concentrations of both particulate (nitrate, ammonium) and gaseous species (nitric acid, ammonia) are updated. The sentence has been changed into

"The concentration of the gaseous precursors are provided by IFS-CB05 and are updated, alongside those of the particulate products (nitrate and ammonium) following the gas/particle partitioning and heterogeneous reaction processed"

P1,L5. “The parameterizations of sources and sinks that have been updated since cycle 45R1 are described” While correct, the sentence is a bit confusing as “are described” refer to ‘the parameterizations’ at the beginning and not ‘since cycle 45R1’

Thanks for the suggestion, the sentence has been changed into

“Here, we detail only the parameterizations of sources and sinks that have been updated since cycle 45R1, as well as recent changes in the configuration used operationally within CAMS”

P1, L9: if you use the IFS acronym it is probably better to say “of IFS” rather than “of the IFS”

Thanks, this has been corrected at various places in the manuscript.

P1, L10-11: “components that are not used operationally will be clearly flagged.” Should I understand that is not the case currently. In that case, should this sentence be highlighted in the abstract?

Thank you, this is a useful suggestion.

P1, L12: a wide range of

Corrected, thank you.

P1, L13: What is meant by an increase in skill?

This has been replaced by “improvement in skill scores”

P2, L15: Not clear what is meant by imbalances

This sentence has been replaced by

“Some problems persist, such as a too high simulated surface concentration of nitrate and organic matter”

L22-26 : This paragraph with many sentences beginning by “Section …” could be improved

This paragraph has been mostly rewritten following the changes in the organization of the paper.

P2, L28: I was not sure what was meant by “bulk–bin scheme ». It should be explained. I don’t think that it the good expression. It seems to be a model approach with a single mode and not a “bin” scheme that represent for me a sectional approach

The “bulk-bin” refers to the fact that IFS-AER is a bulk model for all aerosol species except sea-salt and dust, for which a sectional approach is done, with three bins each. This sentence has been replaced with;

IFS-AER is a bulk aerosol scheme for all species except sea-salt aerosol and desert dust, for which a sectional approach is prefered, with three bins. As such, it is often denoted as a "bulk–bin" scheme; IFS-AER derives from the LOA/LMDZ model (…)

P7, L5: What does the "implementation of a cap" mean?

It means that the emissions are capped if they are above a given maximum value. The sentence has been rewritten into:

"The operational cycle 47R2 doesn't include any update of IFS-AER; the only feature that impacts simulated aerosol fields, besides the upgrade of the meteorological model, is the implementation of a maximum value on primary OM emissions, which is meant to compensate the fact that the emission datasets used underestimate the recent decrease of emissions over China."

Title of Section 4: Primary aerosol sources ?

Corrected, thank you.

P23,L1: remove as in "longer than as simulated"

Corrected, thank you.

P23,L11: Not sure what is meant by ""the very short lifetime … is dominant".

This sentence has been rephrased into:

"The lifetime of sea-salt aerosol is also the lowest of all models, for similar reasons: super coarse sea-salt, with a very short lifetime is much more abundant than the other sea-salt aerosol bins, which show a relatively longer lifetime"

P30, L19: a bias of 2-5 µg/m^3 over Europe does not seem low. Later it says that the bias is negative where the number provided is positive. Is it an underestimation or overestimation

Correct, a negative bias was meant rather than "low bias". This has been corrected here and elsewhere in the manuscript, thank you for the suggestion.

Reviewer #2

Major remarks:

- *I would recommend that the authors consider making some fundamental changes to the manuscript layout. In my view it currently doesn't flow well and this makes it quite hard to read and follow. The model updates sections (Section 3) includes some quite detailed evaluation of the specific updates documented but this is then followed by a further general evaluation in Section 7. Some of the latter still compares Cycle 45r1 with Cycle 47r1 so why not just put all the evaluation aspects together? Could the early sections, detailing model updates not just focus on the difference between the old and new model and then combine all the evaluation together under the Evaluation section. I would also recommend having a Results section which includes current Section 6 and Section 7 as subsections. To me the current layout is a bit disjointed and unclear.*

The authors would like to thank the reviewer for this useful suggestion, which has been adopted. Similar comments from the other reviewers clearly indicate that the overall structure of the manuscript could be improved, and hopefully with this new layout the article will be easier to read through.

- *The model updates are not sufficiently motivated in my view and in many cases the model "improvements" or developments are not reflected in the skill scores. Can the authors motivate the changes in more detail, clearly outlining what the key drivers of the updates were. Are process-based improvements in one part of the model uncovering compensating biases elsewhere within the aerosol scheme? this also should be discussed. Has the original scheme been tuned in any way for instance to give the correct AOD values?*

This is correct; however as this manuscript makes clear, there has been quite a number of model development between cycle 45R1 and cycle 47R1; it is not always easy to pinpoint a single model development responsible for the change in specific skill scores. However, in some cases, it is possible to attribute, at least partly, some improvements to specific model developments: for example, AOD and PM2.5 over Europe and North

America have clearly benefitted from the representation of nitrate and ammonium in cycle 47R1 and from the use of sulfate aerosol production rates provided by IFS-CB05. The Zhang et al. (2014) dry deposition scheme generally has a positive impact on PM2.5 skill scores. Cases of compensating biases almost certainly occur in many instances, but they hard to spot specifically for this kind of rather long and global simulations, given the number of possible sources of errors and simplifications (emissions, production, sink, optical properties). The following sentences have been modified:

- *The description of sulphur cycle in Section 2 would benefit from being described in more detail. The coupling of the aerosol and chemistry schemes is a significant step-change in the complexity of the IFS aerosol scheme and warrants full description. What sources of SO2 are represented and what chemical reactions (gas phase and aqueous phase processes) are represented in CB05, does it include a representation of DMS chemistry for instance. In order to understand the key drivers of the improved evaluation of surface SO2/so4 concentrations is it important to know what processes are represented or not and a reader should not have to go to another reference to get the information needed to understand the results presented here.*

The Sulphur cycle has been described in more details, including aspects that are part of the tropospheric chemistry scheme IFS-CB05. We don't describe the SO2 sources here, as the focus of this paper is rather on the numerical modelling: IFS-AER can be used with any emission datasets, which can then represent different processes. DMS chemistry is included, as well as gas-phase and aqueous phase production of SO_4. The following sentence has been added:

> "The Sulphur chemistry in CB05 is as described in Huijnen et al., (2010). In short, in total 111 Tg SO2 is emitted, which is composed of 97 Tg anthropogenic, 13 Tg volcanic and 1 Tg biomass burning emissions. In addition 38 Tg dimethyl sulfide emissions are applied taken from climatological values, which is oxidized to form SO_2 (37 Tg) and the rest (5.3 Tg) methyl sulfonic acid.
>
> This leads to an annual production of SO_4 of 124 Tg, both through gas-phase oxidation with OH and aqueous-phase oxidation including reactions with H_2O_2 and O_3."

- *Tables 2 and 3 are very confusing and I'm afraid I don't understand them at all. This could be due in part to the Captions perhaps not being complete enough and incorrect labelling used (what config does "IFS" refer to?). I have read and reread the relevant sections but still do not understand why both cycle45r1 and cycle47r1 are included in both tables. I thought coupling to chemistry is included in the latter but not the former so why do we have 4 different simulations of the sulfur cycle? If it is to* *separate out the change in sulphur cycle from coupling to the CB05 alone from the other model upgrades included in cycle 47r1 then this is not at all clear in the text and both main text and table captions need to be improved. Could the 2 tables be merged perhaps to facilitate comparison? The whole section on the coupling and consistent and clear labelling really needs to be improved.*

Thanks for the suggestion: table 2 and 3 have been merged, and the labels reviewed: in particular, in this section, we now consider “standalone” and “coupled” IFS-AER. This section has been almost entirely rewritten so as to make clearer that the focus of this table is really to assess the impact of the coupling of IFS-AER and IFS-CB05 through global budgets and surface concentration evaluation.

- *It would also be nice given the significant impact of the deposition improvements in the cycle47r1 to discuss this generally overlooked part of aerosol modelling, with much focus often being placed on emissions and chemical production etc. The results here highlight the large and important role of more tightly constraining deposition processes in models more generally. It would be nice for the authors to place this work a bit more in the context of the current state of wider aerosol modelling/literature and not just the ECMWF models.*

This is correct, and much more could be presented in terms of evaluation of the Sulphur and Nitrogen deposition fluxes: a lot of work has been done to evaluate IFS-AER in the same way as the HTAP studies (Tan et al. 2018, Vet et al. 2014 for example). A small fraction of this evaluation is shown in this manuscript because it would clearly make it too long and unwieldy. The following two paragraphs have been added to the subsection dedicated to the evaluation of wet deposition against CASTNET data:

“Evaluating and constraining the deposition processes is an indispensable step in the development of IFS-AER and any global aerosol model. There is also a wide interest in the evaluation of Sulfur and Nitrogen (S and N) deposition fluxes, which are generally dominated by wet deposition. Sulfur dioxide and sulfate aerosols impact the acidity of precipitation (Myhre et al. (2017)), while oxidized Nitrogen aerosols and gases (NO3, HNO3 and reduced Nitrogen aerosols and gases (NH4 and NH3) act as powerful plant and microorganisms nutrients when deposited to terrestrial and aquatic ecosystems. On the other hand, excessive input of Nitrogen can lead to eutrophication and loss of ecosystem biodiversity and productivity (Fowler et al. (2015))
Because of this high impact, S and N deposition fluxes are the subject of numerous studies, particularly within the framework of the Task Force on Hemispheric Transport of Air Pollution (HTAP) of the World Meteorological Organization (Tan et al., 2018, Vet et al. 2014). Regionally over Europe, the Eurodelta Trends model intercomparison exercise also focused on S and N deposition fluxes (Theobald et al.

(2019)). IFS-AER coupled with IFS-CB05 is well placed to simulate the gaseous and aerosol components of the S and N deposition fluxes. An exhaustive evaluation of simulated S and N deposition fluxes against the global dataset provided by Vet et al. (2014) has been carried out, but is out of scope for this manuscript. Here, we present an evaluation of the simulated S wet deposition flux"

- *While some areas of the evaluation quantifies the impact of the improvement on model skill scores in others it is more qualitative, and the authors use language such as "x is slightly better than y" or "the skill seems to improve". This I feel detracts from the significance of their findings and from the benefits attributed to these model developments. An attempt should be made to be more quantitative in their language.*

  We agree that this can be improved…The text has been reviewed at various places to add more quantitative statements on model skill scores. Also, a table of mean skill scores has been added so as to provide a quick and easy comparison of the three experiments for the headline skill scores on AOD, PM2.5 and PM10.

- *In parts I find the text a little sloppy and so there are a lot of typographical corrections listed below. Taking a bit more care with the writing would aid both clarity and make the paper easier to follow in places. Figure labelling, I find to be incomplete and doesn't include information on the temporal sampling of model or observed data in many places. For example, in Figure 1 I presume the model data is an annual mean but this isn't clear from the caption. The captions need to be self-explanatory in their own right.*

- *There is insufficient description of the observations used, what time periods do they represent? What temporal sampling was used? This is very important in terms of interpreting the results to understand how representative the comparison is and are you comparing apples with apples! This likely could be better achieved via the restructure of the manuscript recommended above.*

  A new subsection dedicated to the description of the observational datasets used in this manuscript has been inserted. This is now the first subsection of the "evaluation" section.

Minor comments:

P1 L14/15: concentration of sulphate … is improved à concentrations of …. are improved

Corrected, thank you.

P1 L15: imbalances à biases

Corrected, thank you.

P2 L20: Cycling forecasts without data assimilation – presumably it is just the DA of aerosol information that is excluded and the data assimilation of meteorological variables is retained to constrain the simulated meteorology? Please make this clearer in the text.

Yes, the initial conditions of meteorological quantities are provided by an experiment that uses DA, while the initial conditions of aerosols and chemical tracers are the 24h forecasts from the previous cycle. This paragraph has been rewritten as:

“One year of cycling forecasts with 45R1 and 47R1 IFS-AER have been evaluated against an extensive set of ground and remote sensing observational datasets. "Cycling forecasts" refer to experiments which use data assimilation for the meteorological initial conditions, but not for the aerosol and chemical tracers: 24h forecast from the previous cycle are used as initial conditions for these tracers.”

P2 L22-26 : The Section labelling is all incorrect here, as the Introduction is Section 1, needs correcting.

This paragraph has been entirely rewritten as the organization of the manuscript has changed.

P2 L24 : aerosol sources à primary aerosol sources

Corrected, thank you.

P3 L4: All of sea-salt à All of the sea-salt

Corrected, thank you

P3 L6: The use of “etc.” isn’t satisfactory here. You should state clearly what model variables are divided by 4.3 or if all, say “all sea salt properties”

Yes, it is better to be more specific here. This sentence has been expanded into:

“The sea salt aerosol mass mixing ratio as well as the emissions, burden and sink diagnostics need to be divided by a factor of 4.3 to convert to dry mass mixing ratio in order to account for the hygroscopic growth and change in particle density”

Note there is inconsistent use of “sea salt” and “sea-salt” throughout the manuscript. Please make consistent.

Corrected, thank you, we used “sea salt” in the whole manuscript.

P3 Subsection title 2.1.1 : sulfur à Sulfur

Corrected, thank you.

P3 L27: Use of the CAMS_GLOB_ANT emissions versus MACCity, can you more accurately quantify the impact of the different emissions dataset on the subsequent simulated sulfur *cycle. Emissions of SO2 are a big uncertainty in modelled S cycle budget generally and so could play a not insignificant role here? Also in Tables 2 and 3 I presume from the values these are annual mean fluxes but its not clear from the captions.*

It likely has some impact but this has not been evaluated as the IFS is now really using MACCity anymore: these emissions are more or less obsolete now and superseded by the more recent CAMS_GLOB_ANT. Given the wide difference in the simulated surface concentration, it is also likely that this impact is much smaller than that of shifting from a very much simplified conversion scheme to oxidation rates provided by a full tropospheric scheme. A sentence has been added in this paragraph:

"The different sulfur dioxide emissions explain a part of the difference in the budgets and surface concentration plots shown below"

Tables 2 and 3 have been merged, and the label changed so as to make clearer that the values presented are indeed annual mean fluxes and burden.

P3 L29 the chemical conversion rates are globally of the same order of magnitude – please see my comment above on how more detail on the simulation of the S cycle is warranted. Chemical conversion of what to what? Also if the chemical conversion rates are the key drivers of the increase in S lifetime in cycle 47r1 (presumably in Table 2) why is the lifetime of cycle 45r1 in Table 3 similar (~3 days).

The labels have been changed from "chemical conversion" to "chemical loss " (for SO2) and "chemical production" (for SO4). There seems to be a misunderstanding as to what is presented in table 2 (SO2 budgets for 45R1 and 47R1) and 3 (SO4 budgets for 45R1 and 47R1). The two tables have been merged and we tried to clarify the labels. The lifetime of SO2 is much shorter with 45R1 (standalone), indicating a faster production of sulfate aerosol as compared to 47R1 (coupled). The lifetime of sulfate particles on the other hand is not too different between 45R1 (standalone) and 47R1 (coupled).

Table 4: Caption is incomplete and I do not see any comparison or mention of the AeroCom Phase 3 comparison mentioned in the text.

Corrected, thank you.

P6 L5: Put reference in brackets.

Done, and also for other URLs in the text.

Table 5: This table is quite informative but is barely mentioned in the text. Inclusion of the appropriate reference for each cycle would also be good to be included. M86, N12, G14, A16 are undefined. What is meant by Mass Fixer? This also should be explained in the text.

This has been overlooked, thank you for spotting this! The following paragraph has been added:

“IFS uses a semi-implicit semi lagrangian (SL) advection scheme (Hortal (2002)). It is computationally efficient but doesn’tconserve the tracer mass when the flow is convergent or divergent, which is often the case in the presence of orographicfeatures. To compensate for this, mass fixers (MF) are used for green house gases (Agusti-Panareda et al. (2017)), for trace gases (Diamantakis and Flemming (2014)) and for aerosols in IFS-AER since cycle 43R1.”

Also, table 5 is described in more detail in the text: the following have been added:

“This table summarizes the evolution of IFS-AER since 2013, as well as the changes in the emission data sets used and horizontal and vertical resolution.”

The legend of table 5 has been modified to include all missing acronyms.

P9 L2: Monahan86 and Grythe14, why not just label them as M86 and G14 as you do in Table 5. Inconsistent labelling is confusing.

This is a good suggestion, thank you. We replaced the labels in the text.

P9 L5: ofMonahan à of Monahan.

Corrected, thank you.

P9 L5-9 How globally representative are the ocean surface brightness retrievals?

The coverage of the whitecap fraction dataset is really good, even for shorter periods of time (monthly). A sentence has been added:

“As the coverage of the retrieved whitecap fraction data set is very good, the sample size is very large, which makes the fit quite robust.”

P9 L24: Similarly to à Similar to

Corrected, thank you.

P9 L29: How was the evaluation carried out? Temporal frequency of observations and model (again not stated in caption of Figure 5 but it looks to be weekly?)? How representative is this comparison? The sea salt contribution to the total AOD will be maximised in local wintertime, and so exhibits a clear seasonal cycle, has this been assessed? The MAN network could otherwise contain contributions from secondary sources of sulfate aerosol from DMS and other biogenic sources, looking at the seasonal cycle could help discriminate between the various sources.

For MAN, the observational frequency is daily (once a day): as model data is archived 3 hourly, we used the closed 3 hourly simulated value to compare against MAN data. Similarly for AERONET, the closest simulated value in time (3h precision) and space is used, and then the skill scores are averaged weekly. More information on the time sampling has been added. We are aware that both MAN and the selected AERONET stations can include also contributions from other species than sea salt aerosol, as written: "Total AOD provided by MAN cruises generally, but not always, consist mostly of sea salt aerosol optical depth. Occasionally, dust or biomass burning plumes can also have an impact, but this impacts a minority of the measurements.". This reduces the significance of the comparison; which is why we tried to look at this from different angles: MAN, AERONET, merged AOD product from FMI as well as surface concentration when available. The fact that all of these evaluation are consistent in showing an improvement in skill scores with the A16 scheme gives us some confidence in its added value in IFS-AER. We agree however that not too much weight should be put to a single evaluation of sea salt aerosol quantities, taken in isolation of the others.

P10 L2 AEROCE/SEAREX programme – include appropriate citation

A citation to Savoie et al. (2002) has been added.

P10 L8: is slightly improved -> can you be more quantitative

A quantitative assessment is given:

"The root mean square error (RMSE) is slightly improved with CY47R1, decreasing from 0.115 to 0.111"

Table 6 and 7: It would be good to include the diameter ranges below the bin labels.

Thanks for this useful suggestion.

P14 L8 and Figure 6: is this an annual total?

Yes, the first sentence of the corresponding paragraph has been changed into

The 2017 annual total (sum of all bins) dust emissions with the two emission schemes is shown in Figure 6.

The Figure label has also been modified to clarify this.

P14 L14: The skill of the simulated dust seems to improve -> please be more quantitative

This sentence has been expanded into:

"The skill of the simulated dust seems to improve with the new scheme, as RMSE is generally decreased and sometimes nearly halved, such as in August-September 2017 with a value of 0.1 instead of 0.2. The spatial correlation is also generally improved and stands above 0.9 most of the time with CY47R1 while it is frequently below this value with CY45R1."

P14 L21: IFS-AER – which version? Please cf with reference to IFS-AER on L22 (same page)

Corrected, thank you. This refered to CY45R1 and CY47R1 for the second instance.

P14 L21 producign à producing

Corrected, thank you.

Figure 7: It would be nice to see some uncertainty bounds or even a standard deviation of the observations (dust being highly variable in space and time) on these plots.

The temporal variability is indeed high, which is why we chose a weekly AOD plot instead of daily AOD. However, the verification software we currently used doesn’t allow to add such information, even though it would be useful to have.

P18 L3: have been brought to à have been added?

This is better, thank you for the suggestion.

P19 L14/15: why bold? Also I don’t really understand how something can be implemented in a cycle but is not operational? Do the cycle numbers and revisions not refer to an operational configuration?

The mention “which is not yet used operationally for technical reasons, but is used in operational cycle 47R3” is in bold because it refers to a development that is part of the operational cycle 47R1 but which is not used. It is used operationally in CY47R3 (October 2021) and later. There are several dry and wet deposition options that are implemented in the operational cycle, but only one of them is used. Because of a technical bug, the newly implemented dry and wet deposition options have been used operationally in cycle 47R3 instead of 47R1.

P20 L14: This sulphur à The sulphur . Also sulphur and sulfur are both used in the text.

Corrected, thank you. We now only use “Sulphur” in the text.

P21 L2: CASTNET – include appropriate reference?

References for the CASTNET and EMEP have been added.

P23 L1: Where is the budget for Cycle 45r1 presented?

They have been included in the 2019 GMD paper, and we felt it was too much of a repetition to include them in this paper again.

P23 L9/L18: “20” missing unit

Corrected, thank you.

P23 L29: Are the AeroCom values for the year 2017, if not this would easily explain differences in emissions?

The AeroCom values are for the year 2010; besides, the AeroCom simulations used different anthropogenic emission inputs (CMIP6 vs CAMS_GLOB_ANT). As such, the two are not completely comparable. The following two sentences have been added:

"The AeroCom median refers to the 2010 experiment. Because the AeroCom experiments are for 2010 and used a different set of emission inputs, the median is not fully comparable to values provided by IFS-AER simulations of the year 2017. However, they give an indication of how IFS-AER broadly compared to other global aerosol models."

P25 L6: deserts a à deserts

Corrected, thank you.

Figure 12: Highest and lowest values use the same colour which is a bit confusing.

The plots have been redrawn in order to avoid this.

P27 L14/15: what drives the simulated peaks in AOD? Are SO2 emissions from fire included?

Yes, SO2 emission from fires are included, but the spikes in the simulated AOD are caused by organic matter AOD (from biomass burning sources). It has been mentioned in the text:

"In July and August 2017, large fires in the U.S. and Canada provoked spikes in simulated and observed AOD (consisting mainly in organic matter AOD)"

P27 L29 : positiv à positive

Corrected, thank you.

P29 L6: significantly over à significantly improved

Corrected, thank you.

Figure 14 caption: regionallevel à regional level

Corrected, thank you.

P30 L8: probably don't always hold true à isn't true in all instances

This paragraph has been rewritten and moved to another subsection ("observations used")

Figure 16: While the evaluation of dust deposition is qualitative at best, it does look like the model deposits most of its dust too close to the African coastline with not enough extending westward over the Atlantic.

Yes, the transatlantic transport of dust is probably underestimated. The following sentence has been added:

“However, it seems that deposition is overestimated close to the African coastline, and underestimated in the Western Atlantic, a sign that transatlantic transport of dust is possibly underestimated by IFS-AER.”

P30 L24: 2-3 ug/m3 than à 2-3 ug/m3 more than

Corrected, thank you.

P31 L2: biomassburning à biomass burning

Corrected, thank you.

P31 L4/5 : Again the improvements associated with the NEWDEP changes are very interesting. Are the new deposition changes offsetting the increase in AOD and PM2.5 associated with the biomass burning emission height change? Presumably if more particles are emitted higher up away from BL processes and sedimentation processes. Do you see a shift from dry to wet deposition between 45R1 and 47R1? Also the NEWDEP changes seem to impact some aerosol species (eg biomass burning) more than others (eg nitrate), can the authors offer a suggestion as to why this is the case?

This is a very good question ;-) The NEWDEP changes generally increased dry deposition, particularly for fine aerosols, as shown by the budgets (table 11 in the first version, table 10 in the new version of the manuscript). Why this is so is partly explained by the bin-like approach of the Zhang and He (2014) dry deposition scheme, as compared to the continuous dependency on particle size of the Zhang et al. (2001) dry deposition scheme, as shown in Figure 9. Although wet deposition is still much larger than dry deposition for fine aerosols, PM2.5 is clearly driven primarily by dry deposition (and emissions of course) rather than by wet deposition. The surface concentration of OM is significantly reduced by the NEWDEP dry deposition algorithm. Why this is less so for nitrate is I think consequence of what the impact of NEWDEP on sulphate: Increased dry deposition with NEWDEP reduces the surface sulfate concentration, thus allowing for a higher production of nitrate close to surface.

P31 L7-10: Improvements in PM biases over China, can this be linked back to improved S cycle via coupling of aerosol scheme to CB05 chemistry?

Yes, this is clearly one of the dominant factor in the improvement there. The following sentence has been added:

“The improvement with CY47R1 is largely explained by the improved representation of the sulphur cycle in CY47R1 associated with the coupling to IFS-CB05.”

P31 L25: obsreved à observed

Corrected, thank you.

P31 L27: in which the same nitrate scheme as IFS-AER has been adaptedà which has implemented an adapted version of the IFS-AER nitrate scheme

Much better, thank you for the suggestion.

Figure 15: is only half the possible range of FGE covered by the colorbar used here? There seems to be a large underestimation of AOD over the Maritime Continent that persists through all model versions.

The whole range of the colorbar values is not used in the FGE plot, but this is also because a common colorbar is used for the AOD values and for the FGE. Yes, there is a clear underestimation of AOD over the maritime continent. Some contributing factors could be the fire emissions over Indonesia, and also too little biogenic secondary organic aerosols there.

Figure 18 caption : Please note the different scale between the two panels – but the scales appear to be the same?

This was an error, now corrected, thank you.

Figure 19: I can't make out the observations in black circles.

Figures 18, 19 and 20 have been redrawn to make CASTNET/EMEP observations more visible.

Figure 20: what time period is used for the observations?

This is the 2017 yearly average, as for the simulated value. The label has been modified

Figure 20 caption: should this be OM in PM2.5 and not surface ammonium concentration?

Corrected, thank you.

Reviewer #3

- Please make a harmonisation in terms of model simulations used. I believe it would be useful if all three model versions 45R1 and 47R1 and 47R1newdep should be shown for all evaluations and tables. I think also that for some evaluation the period February to December 2017 was used, and for others Jan-Dec 2017. Not sure this change in base time period is useful. Even though results for 45R1 are probably (? I did not check ?) already in Remy 2019, they should also be shown here.

We changed the evaluation period to be 2017 (January to December) in the whole manuscript: as you pointed out, using January-December or February-December 2017 was confusing and misleading. Because this article focuses on 47R1, we chose to show 45R1 results selectively instead of systematically. The paper is already long, and we thought it would be burdening really too much to include 45R1 in all plots and tables.

- The model changes should be described first and then evaluation results could be discussed altogether in subsequent chapters.

Thank you for the suggestion. Following your remark, shared by other reviewers, the structure of the paper has been changed to better separated the model description and the model evaluation parts.

- The statistical data are presented in four different ways: as maps, as time series, as tables, as histograms. It is not clear why the "style of presentation" changes in the course of the manuscript. It would be good to have the same statistics, eg bias, rmse, r, and mnmb available for all evaluations. This could be a few overview tables or more complete annotations in the figures. This would give substance to the often vague statements in the text on quality. Please make sure statistics are available for all three model versions 45R1 and 47R1 and 47R1newdep.

In the new version of the manuscript, we tried to rationalize the metrics used and focus on Fractional Gross Error, except for a few plots (AOD for dusty AERONET stations) instead of on RMSE. Also, a table that summarizes MNMB and FGE of simulated AOD, PM2.5 and PM10 for 45R1, 47R1 and 47R1_NEWDEP has been included. The labels of the plots and tables have been reviewed to add more information.

- from table 5: It looks like 45R1 and 47R1 are run on quite different vertical resolution. Shouldnt that be a major factor in all aerosol budgets? This is not discussed as far as I can see it. If the experiments have been made on different vertical resolution, then the changes in budgets are not just because of the changes parameterisations. This would be interesting to understand for the budgets and the deposition evaluation.

This is true of the operational configuration: there are 60 levels for CY45R1 (June 2018 to July 2019) and 137 levels for CY47R1 (October 2020 to May 2021). However, in the simulations that have been carried out and evaluated in this paper, we have been careful to use the same horizontal (Tl511), vertical (137 levels) and emission inputs (CAMS_GLOB_ANT) between the runs. In order to clarify this, a sentence has been added at the end of the paragraph that introduces the forecast only simulations:

“In order to assess the model skill independently of resolution and emission inputs, the three simulations 45R1, 47R1 and 47R1_NEWDEP used the same horizontal and vertical resolution, and the same emission inputs”

- In general, I am not clear about whether the integration to the IFS CB05 is activated in all 47R1 experiments with IFS-AER? Does mentioning 47R1 and IFS-AER mean (for the experiments shown in this study) that IFS-CB05 is used for all gas phase chemistry?

The coupling is used by default in all CY47R1 simulations. However, IFS-AER can still be run standalone if the user wishes so. All the CY47R1 simulations presented in the manuscript are coupled to IFS-CB05. The following sentence has been added to the same paragraph that describes the experiments:

“The two CY47R1 and CY47R1_NEWDEP experiments use IFS-AER coupled to IFS-CB05.”

specific comments:

- table 5 - I think this table comes a bit late.

The manuscript has been re-organized, and the section 2 has been shortened significantly: table 5 now comes earlier

- table 2+3: Should be simply combined to one sulfur cycle table. Its a bit confusing this way with two tables.

The two tables have been merged, thank you for the suggestion. The labels have also been updated to make this table clearer.

- table 2: SO2 budget - Its a bit counterintuitive that life time goes up when wet deposition as a process is added. Why did dry deposition go down so much in 47R1?

This is because the conversion is very different and clearly generally slower when IFS-AER is coupled to IFS-CB05.

p9 l2: typo “Grythe14 Grythe”

Corrected, thank you.

- table 4: the figure caption is incomplete. What is in brackets? Which cycle is shown? maybe show all?

Corrected, thank you. This is CY47R1; this info has been added in the table legend.

- p7 l11: 46R1 is discussed but what about 45R1?

The detailed operational configuration of 45R1 is presented in the 2019 paper; a summary is available in Table 5. 46R1 was mentioned because it was not described in the 2019 paper.

- table 5 I think the experiments used in this paper: 45R1 and 47R1 and 47R1newdep without data assimilation should be included in this table to clarify what is used.

We prefer not to include them here, as this table lists only the operational configuration of each cycle: if we include specific experiments, it maybe confuse CAMS users if they refer to this table to know the specifics of the IFS-AER configuration.

-p2 l20 : one year of cycling without data assimilation : I think the missing data assimilation should be commented more . How different is the model as compared to the operational model with data assimilation?

It is really quite different, for AOD, PM2.5 and surface concentration of sulphate, nitrate and other species. However in this paper we prefer to not discuss results with data assimilation, as it focuses on the description and evaluation of IFS-AER, but not of CAMS products. Detailed evaluation of the global CAMS products (including aerosol products) can be found in https://atmosphere.copernicus.eu/publications. A sentence has been added into Section 3 with a link to this page:

"The operational global CAMS products are routinely evaluated against a variety of observational data sets. The quarterly evaluation reports are available online and can be consulted at <\url{https://atmosphere.copernicus.eu/publications}>."

-2.1.1 sulfur => Sulfur

Corrected, thank you.

-p3 l29: sinks=> sink

Corrected, thank you.

- table 6: Why is the life time changing so much for a given bin and different source functions? Each bin has one size and density. Dry and wet removal should be roughly the same, or?

The lifetime of sea-salt aerosol depends a lot on particle size: sedimentation is a very efficient sink for super coarse particles, and dry deposition depends a lot on particle size: as shown by Figure 9, the dry deposition velocity increases very quickly as a function of particle size above 1-2 micron diameter. You are correct that wet deposition doesn't change too much for each of the three bins. The following sentence has been added:

"The lifetime of sea-salt aerosol decreases for larger particles, because sedimentation, applied only to bin 3, is an effective sink, and because the simulated dry deposition velocity increases with particle size for particles above 1 micron diameter"

- figure 3: What is the color in the plots? It looks like all dots are plotted…

The colors indicate the density of the scatterplot: yellow is denser, dark blue is less dense.

- figure 7: typo at end

Corrected, thank you.

- p22 l15: Budgets are shown for Feb-Dec but then table 11 says Jan-Dec 2017, what is used ? Why not using Jan-Dec throughout? Is the spinup really needed?

We followed your advice and showed all plots and tables for the January-December 2017 period. In a forecast only period, the spinup is only a few days and shouldn’t impact much the values and plots presented.