Dear referee#2 ,
thank you very much for your review of our manuscript GMD-2019-07. Please find our replys to your comments below. In the following, referee comments are given in italics, our replies are in normal font, and text passages which we included in the text, are in bold.

This manuscript explores whether the source apportionment of surface ozone would be affected by model resolution. It performed model simulations with the different resolution of the model itself and the emission inventories. The difference in the source apportionment using a self-consistent tagging method is attributed to the model resolution and emission inventory resolution. The topic itself and the self-consistent tagging method are interesting.

Reply: We thank referee#2 for this summary and honouring our work with the self-consitent tagging method.

However, the analyses presented in the manuscript are too useful;
Reply: We do not understand this comment.

the discussion and conclusions are not insightful (or not having any new results as pointed out by Anonymous Reviewer#1). I'd suggest the following items to improve the manuscript.
Reply: First of all we thank referee#2 for the ideas on how to improve the manuscript which we comment in detail below. As already discussed in our reply the referee#1 we think that we provide new results, because, at least to our knowledge, the impact of the model resolution (and other technical factors) on the results of source apportionment methods has not been investigated in detail. Such an investigation, however, is important for two reasons:

- To investigate how robust the source apportionment results from global models are on the regional scale, and
- to estimate the range of uncertainties of source apportionment caused only by technical limitations of the models and emission inventories.

Even tough our results are only valid for a specific model, they provide new insights about possible ranges on model caused uncertainties. Such results are important for the community involved in source apportionment methods, both on the global and the regional scale.

Finally, we would like to remark that publications in GMD are not primarily about presenting new scientific results. Publications in GMD are mainly to document model developments, document experimental set-ups of model simulations, document evaluation of model systems, present model evaluation strategies and to present technical analyses of model systems. We think that our manuscript documents the influence of model and emission inventory resolutions on source attribution results. This is clearly important to asses source

apportionment results and their related uncertainties, also for other model systems.

To underline the importance of our findings to other modelling communities we largely rewrote the conclusion section. The new part reads:

Apart from many model specific findings of this study, its results have important implications for other modelling studies and modellers applying source apportionment methods. These implications are:

- **First, our study shows that average continental contributions of anthropogenic emissions are quite robust with respect to the used model and the used model resolution. This means that global models at coarse resolution can be used to perform ozone source apportionment in this global context.**
- **Second, our results also show that on the regional scale, the differences either caused by different models, but also by model resolution are much larger. These effects arise mainly near hotspot regions like the Po Valley or near major shipping routes in the Mediterranean Sea. However, especially in these areas, contribution analyses of anthropogenic emissions are very important and spurious effects, such as artificially increased ozone levels and contributions caused by the coarse resolution of models and or emission inventories should be avoided. Hence, for regional analyses fine resolved models and emission inventories are required.**
- **Third, our results clearly indicate how large the spread between models with respect to STE is. The importance of stratospheric ozone, both in the global and the regional model, corroborates the necessity of tracing the contributions of stratospheric ozone to ground level ozone explicitly by the source apportionment methods. However, only few currently available methods used on the regional scale account for this process.**

1. better defining the differences between simulations/models, be specific about what processes causing the variations in source apportionment. Here are just a few examples to improve.
Reply: As discussed in detail below we think that we are discussing a lot of processes causing these variations in detail. We agree that some explanations could be improved (see below). To better define the model and simulation differences we revised the manuscript accordingly and added the Section 2.2 in which we discuss the different simulations and the motivation for performing these simulation in more detail.

(a) Meteorological inputs such as temperature and light are different in some simulations, which would result in different biogenic emissions in methods of

'on-line calculated' and 'calculated by EMAC.'

Reply: No: As stated on p5l32ff of the original manuscript MECO(n) is applied in the so called quasi-chemistry transport model mode (QCTM-mode). In this mode the coupling between chemistry and dynamics is disconnected and each model instance simulated the same meteorology in all simulations. Of course, the dynamics differs between the different model instances due to different resolutions and/or physical parametrizations, which leads to differences in the biogenic emissions. We have discussed this issue on p3l4ff (of the original manuscript). For this reason the simulation *EBIO* is performed to investigate the impact of different biogenic emissions.

We added a note about QCTM in Sect. 2.2:
In this mode chemistry and dynamics are decoupled to increase the signal-to-noise ratio for small chemical perturbations. **This means, that even tough the emissions differ in the different simulation each model instance (EMAC, CM50 and CM12) simulated the same dynamics in all simulations. The dynamics between EMAC, CM50 and CM12, however, differs due to different resolution and physical parameterisations.**

Further, we added a longer description on the motivation of the *EBIO* simulation:
In the simulation *EBIO* the biogenic C_5H_8 and soil-NO_x emission as calculated by EMAC are transformed down and applied in CM50. By comparing the results from CM50 of *REF* and *EBIO* the effect of the different biogenic emissions can be analysed. These differences of the biogenic emissions are due to differences in the simulated meteorology between EMAC and CM50.

(b) Same anthropogenic emissions in different resolution might result in the same total emission but large regional differences. How do these emissions differ?

Reply: The coarse resolution of the emissions leads to a dilution of emissions over larger areas. Please see Fig S1 showing the MACCity land transport emissions in EMAC and in CM50. This figure is also added to the revised Supplement. Further, we added tables with the total emissions of the different simulations to the Supplement (Table S2-S10 in the new Supplement). To investigate the impact of the emission resolution onto the results the simulation *ET42* was performed.

(c) What are actually causing the differences in STE flux in the coarse vs. fine resolution model? Could it be related to on-line vs. off-line meteorology/convections and/or temporal and horizontal averaging of meteorological inputs (just some examples I am familiar with, like in Yu et al. (2018) and Hu et al. (2017); certainly, many other literature on this topic are available)? The

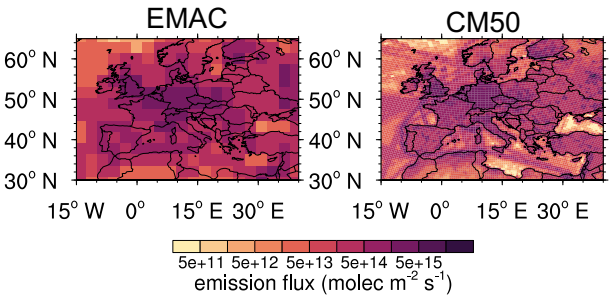

Figure S1: Annual averaged emissions flux (in molec m^{-2} s^{-1}) of NOx due to all anthropogenic emission sources (land transport, anthropogenic non-traffic, shipping; *REF* simulation) for EMAC and CM50.

contribution from downward transport seems to be the largest differences among models, and it should be quite interesting to explore.

Reply: Indeed, the largest differences between EMAC and CM50 are the differences of the STE. EMAC and COSMO/MESSy are chemistry-climate models, no chemistry transport models. Hofmann et al. (2016) already investigated in detail differences of the STE between EMAC and COSMO/MESSy, therefore we don't want to discuss this topic in detail again. Generally, the finer resolution of COSMO/MESSy leads to a better representation of the physical processes of individual STE events. However, in our manuscript we do not focus on individual events but rather on multi-year average values. For these multi-year average values the increased contribution of ozone from stratospheric origin is mainly confined to the planetary boundary layer. The reason for this is more efficient vertical mixing in COSMO/MESSy, partly caused by more vigorous convection and by an too unstable boundary layer during night. Taking the biases compared to observations into account this vertical mixing in COSMO/MESSy seems to be too strong, which indicates that the larger contribution of stratospheric ozone (and also for the categories aviation, lightning and N2O) is an artefact of this too strong vertical mixing. As discussed below we added a new Sect. 3 including a model evaluation to the manuscript. Further, we discuss the reason for the STE difference in more detail in the revised manuscript (see various changes in Sect. 4)

(d) It looks like the total lightning NOx emissions are the same across simulations, do their also have the same 3D distribution?
Reply: Yes. Over all simulations and over all model instances the same lightning-NO_x emissions are calculated. These are the emissions calculated by EMAC which are transformed during runtime from the EMAC grid onto the grid of CM50/CM12. The procedure is described in the model description section, but we rephrased the description to make it more clear. The new sentence is:

The lightning NO_x emissions are calculated only in EMAC using a

parametrization based on Price and Rind (1992), which is scaled to a global nitrogen oxide emission rate of $\approx$ 5 Tg(N) a^{-1} from flashes. In CM50 and CM12 we use the emissions from EMAC (i.e. with same geographical, vertical and temporal distribution), which are transformed on-line onto the grids of CM50 and CM12, respectively.

'Inter-model differences' should be better defined and documented and can provide insights on the calculated contributions. Specific discussion of these processes rather than vaguely saying because of the resolution would make this paper more useful.

Reply: The manuscript is about discussing the processes and other possible explanations for the differences between the different model results. Some examples are (page and line number refer to the original manuscript):

- p9l5ff [..] Due to increased vertical mixing in CM50 compared to EMAC ozone which is produced in the upper troposphere is transported downward more efficiently. [..]
- p10l12ff [..] mainly caused by a decreased net ozone production and a stronger vertical mixing in CM50 compared to EMAC. [..]
- p11l3ff [..] As the analyses of the *ET42* simulation results shows, the coarse resolution leads to an artificial increase of P_O3 which in turn leads to an increase [..]
- p11l10ff [..] On the coarse EMAC grid most parts of Southern Italy are considered as sea, affecting especially the calculation of dry deposition in EMAC, as dry deposition of ozone is lower over sea as over land.[..]
- p11l24ff [..] Especially in Southern Germany this is mainly caused by the better resolved topography and larger contributions of stratospheric ozone [..]
- p11l30ff [..] in Western Germany CM12 simulates a larger contribution of the CH_4-category to ozone compared to CM50, which is consistent with the larger tropospheric oxidation capacity in CM12 compared to CM50 (Mertens et al., 2016). [..]

We used the term 'inter-model differences' in some parts of the original manuscript to refer to the differences which we discussed before. In some parts we also referred to previous findings of Mertens et al. (2016). We rephrased these parts to be more precise. As an example we added the following note in the newly added Sect. 2.2:
Differences between the results of the EMAC and CM50 (and CM12) can be attributed to different effects: First, the dynamical core and physical parametrizations between EMAC and COSMO/MESSy differ, second the resolution of these models differs and third EMAC

and COSMO/MESSy calculate different soil-NO_x and biogenic C_5H_8 emissions. The latter due to the meteorology dependence and due to different soil types in EMAC and COSMO/MESSy.

Similarly, we added more detailed explanations in Sect. 3 and Sect. 4: An example from Sect. 3:
As already noted by Mertens et al. (2016), CM50 exhibits a larger positive ozone bias compared to EMAC. This bias is mainly caused by a more efficient vertical mixing in COSMO, as well as by a less stable boundary layer during night. The latter is a common problem of many models leading to diurnal cycles with too large ozone values during night, which results in an overall ozone bias (e.g. Travis and Jacob, 2019). The coarser resolution of the emissions (*ET42*) as well as the different biogenic emissions (*EBIO*) between EMAC and CM50 contribute only partly to the bias of CM50 compared to EMAC. The CM50 ozone bias is larger in *ET42* and *EBIO*. The pattern of the ground level ozone mixing ratio bias of CM50 compared to EMAC is similar for all simulations (see Fig. 3). Generally, CM50 has a positive ozone bias compared to EMAC over most parts of Europe.

Further, we add some more discussion about the differences of the stratospheric contribution between CM50 and CM12. These differences can mainly be attributed to stronger vertical mixing caused by stronger updraft and downdraft massfluxes in CM12 compared to CM50.

2. the terms used in the manuscript are very confusing for readers from outside the MESSy model community, particularly when referring to the specific simulation. For example, CM50 is used to compare with EMAC, while one refers to the resolution of 50km of one model; the other refers to a different model. ET42 refers to 'the MACCity emissions are transformed to the coarse grid of EMAC (T42), to investigate the impact of the resolution of the emission inventory.', but it sounds like it is done by the COSMO model only, so do all the REF, EBIO, EVEU simulations. Table 2 seems to suggest that EMAC also has those four simulations. Table 1 is not useful in the context of this manuscript but just adds confusions by adding a bunch of acronyms. This manuscript should not be 'read very much like a technical report for MESSy users' as pointed by the other reviewer. Readability should be improved.

Reply: We have th feeling that some of the confusion is caused by a missunderstanding of the concept of the MECO(n) model system. However, as the concept is explained in detail in a series of 5 different papers cited in Sect. 2 we wanted to recap only the basic concept of MECO(n). Obviously this basic recap was too short. Therefore we added a slightly longer description of MECO(n). This new part reads:
We apply the MECO(n) model system, which couples the global chemistry-climate model EMAC during runtime (i.e. on-line) with

the regional chemistry-climate model COSMO-CLM/MESSy (Kerkweg and Jöckel, 2012b). Both models, EMAC and COSMO-CLM/MESSy, calculate the physical and chemical processes in the atmosphere and their interactions with oceans, land and human influences. They use the second version of the Modular Earth Submodel System (MESSy2) to link multi-institutional computer codes (Jöckel et al., 2010). The core atmospheric model of EMAC is the 5th generation European Centre Hamburg general circulation model (ECHAM5, Roeckner et al., 2006). The core atmospheric model of COSMO-CLM/MESSy is the COSMO-CLM model (Rockel et al., 2008), a regional atmospheric climate model jointly further developed by the CLM-Community based on the COSMO model. In the model systems acronym 'n' denotes the number of COSMO-CLM/MESSy instances nested into the global model framework. The initial and boundary conditions, which are required for each of these nested regional model instances, are provided by the next coarser resolved model instance. This model instance can either be EMAC or COSMO-CLM/MESSy. Due to the on-line coupling the boundary conditions for the regional model instances can be provided at every time step of the driving model instance. This especially important to resolve short term variations of chemically active species. As EMAC and COSMO-CLM/MESSy calculate both, atmospheric dynamics and composition, the meteorological and chemical boundary conditions are as consistent as possible. In addition, the same chemical solver and kinetic mechanism is applied, leading to highly consistent chemical boundary conditions. Therefore, there is no need of lumping (i.e. treading different chemical species with similar chemical formula as one species), scaling boundary conditions for specific chemical species or taking boundary conditions from different models. More details about the MECO(n) model system are presented in a set of publications including a chemical and meteorological evaluation (Kerkweg and Jöckel, 2012a,b; Hofmann et al., 2012; Mertens et al., 2016; Kerkweg et al., 2018). The set-up of the simulation applied in the present study is very similar to that described by Mertens et al. (2016). Therefore, we present only the most important details of the model set-up. The complete namelist set-up is part of the Supplement.

It is important to understand that in every model simulation different instances of the MECO(n) model run at the **same** time and share necessary boundary and initial fields via MPI communication. For the applied MECO(2) set-up the running model instances are: EMAC, COSMO/MESSy with 50 km resolution (named COSMO(50km)/MESSy) and COSMO/MESSy with 12 km resolution (named (COSMO(12km)/MESSy)). These terms where introduced by Hofmann et al. (2012) and to ease readability the short terms CM50 and CM12 were introduced by Mertens et al. (2016). We don't want to add confusion by introducing new terms and therefore stick to these previously introduced abbre-

viations. We think that this is the best way of having a clearly defined model system. However, we are open for concrete suggestions for an improved naming.

To make clear thatin MECO(n) different model **instances** run at the same time (and we do not perform simulations with different models) we used the term troughout the revised manuscript.

Using the MECO(2) system (EMAC → CM50 → CM12) we performed different simulations (*REF*, *EVEU*, *EBIO* and *ET42*) and compare the results of all three model instances. To differentiate between model instances and simulation names, the simulation names are written in italics throughout the manuscript. As shown in Table 2, EMAC is running in all simulations, but with the same set-up including the same emissions. To make this more clear we added a new subsection called 'investigation concept' (Sect. 2.2).

With respect to Table 1 we do not agree with referee#2. We think this table is the best and shortest way of showing the model set-up, which is (to our opinion) very important in terms of reproducibility. We list the name of the individual submodels as they are published in peer reviewed literature under these names. For people not familiar with these submodels we have a short description, stating the physical/chemical process or the diagnostic provided by this submodel as well as a reference describing the individual submodel in detail.

3. this paper could benefit from a section of model evaluation by adding comparisons with observations. This way could suggest which simulations are 'in practice' better and if the model simulations are actually realistic.

Reply: We added a basic section of model evaluation and compare the performance of the individual model instances for the different simulations with observations (new Sect. 3). For this, we use ground based station measurements as well as ozone sonde measurements. This should give an impression of the overall model performance. However, from this model evaluation it is not possible to evaluate ozone contributions as these are pure model diagnostics.

4. the metrics used to quantify simulation difference: this manuscript mostly uses the average concentration of ozone and relative contribution of a specific source. These tend only to show minimal differences among simulations; even though the manuscript claims 'up to 20%' in the calculated contribution of transport emissions, the absolute amount is small. One way to improve is looking at the probability distribution of concentrations or contributions, which could be much more useful to examine differences in model chemical pathways and for specific air pollution episodes, i.e., examples like in Fiore et al. (2002) and Yu et al. (2016).

Reply: We do not fully agree with this comment. In Fig. 3 and Fig. 7 we show box-whisker plots which indicate the range of the simulated values (e.g.

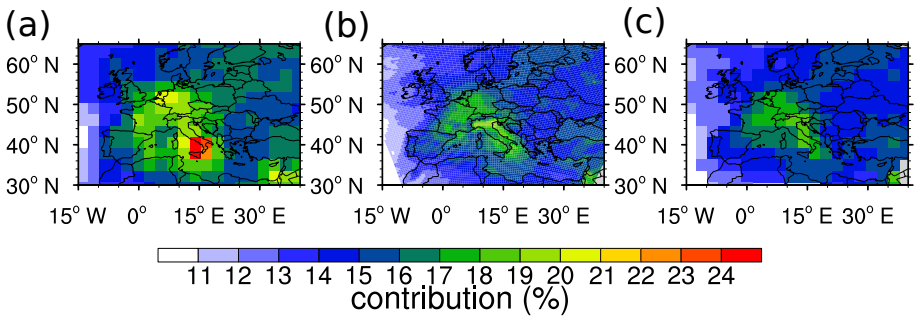

Figure S2: 95th percentile of the contribution of O_3^{tra} to ground level O_3 (for JJA between 9–18 UTC) for **(a)** EMAC, **(b)** CM50 and **(c)** CM50 transformed onto the EMAC grid ($CM50_E$).

range, 25th and 75th percentile, mean and median). Of course we could also have chosen PDFs instead, but they offer similar information. Further, we also discuss differences of the 95th percentile of the contributions of land transport. The differences of these extreme values are of course larger than the average differences.

We agree that the differences for these averaged values show only small differences, but the focus of your analysis is on time scales on which global models (e.g. multi-year averages) and not on the scale of individual pollution events. Therefore, we prefer to stick to the applied metrics. Further, we are not aware where we claim 'up to 20%' in the calculated contributions of transport emissions', as we claim a differences of up to 20 % between the simulated contribution of the different models/model set-ups. These differences are simulated around the Naples region, were the relative contributions between EMAC (17 %) and CM50 (13 %) differ. These relative contributions refer to absolute contributions of 3 to 4 nmol mol^{-1}. For the 95th percentile (see Fig. S2) of the relative contribution of O_3^{tra} these difference increase to around 6 percentage – points. We added this figure to the Supplement (Fig S6).

We are looking forward to your reply,

Mariano Mertens
(on behalf of all co-authors)

References

Hofmann, C., Kerkweg, A., Wernli, H., and Jöckel, P.: The 1-way on-line coupled atmospheric chemistry model system MECO(n) Part 3: Meteorological evaluation of the on-line coupled system,

Geosci. Model Dev., 5, 129–147, doi:10.5194/gmd-5-129-2012, URL http://www.geosci-model-dev.net/5/129/2012/, 2012.

Hofmann, C., Kerkweg, A., Hoor, P., and Jöckel, P.: Stratosphere-troposphere exchange in the vicinity of a tropopause fold, Atmospheric Chemistry and Physics Discussions, pp. 1–26, doi:10.5194/acp-2015-949, URL https://doi.org/10.5194/acp-2015-949, 2016.

Jöckel, P., Kerkweg, A., Pozzer, A., Sander, R., Tost, H., Riede, H., Baumgaertner, A., Gromov, S., and Kern, B.: Development cycle 2 of the Modular Earth Submodel System (MESSy2), Geosci. Model Dev., 3, 717–752, doi:10.5194/gmd-3-717-2010, URL http://www.geosci-model-dev.net/3/717/2010/, 2010.

Kerkweg, A. and Jöckel, P.: The 1-way on-line coupled atmospheric chemistry model system MECO(n) Part 1: Description of the limited-area atmospheric chemistry model COSMO/MESSy, Geosci. Model Dev., 5, 87–110, doi:10.5194/gmd-5-87-2012, URL http://www.geosci-model-dev.net/5/87/2012/, 2012a.

Kerkweg, A. and Jöckel, P.: The 1-way on-line coupled atmospheric chemistry model system MECO(n) - Part 2: On-line coupling with the Multi-Model-Driver (MMD), Geosci. Model Dev., 5, 111–128, doi:10.5194/gmd-5-111-2012, URL http://www.geosci-model-dev.net/5/111/2012/, 2012b.

Kerkweg, A., Hofmann, C., Jöckel, P., Mertens, M., and Pante, G.: The on-line coupled atmospheric chemistry model system MECO(n) – Part 5: Expanding the Multi-Model-Driver (MMD v2.0) for 2-way data exchange including data interpolation via GRID (v1.0), Geoscientific Model Development, 11, 1059–1076, doi:10.5194/gmd-11-1059-2018, URL https://www.geosci-model-dev.net/11/1059/2018/, 2018.

Mertens, M., Kerkweg, A., Jöckel, P., Tost, H., and Hofmann, C.: The 1-way on-line coupled model system MECO(n) – Part 4: Chemical evaluation (based on MESSy v2.52), Geoscientific Model Development, 9, 3545–3567, doi:10.5194/gmd-9-3545-2016, URL http://www.geosci-model-dev.net/9/3545/2016/, 2016.

Price, C. and Rind, D.: A simple lightning parameterization for calculating global lightning distributions, J. Geophys. Res. Atmos., 97, 9919–9933, doi:10.1029/92JD00719, URL http://dx.doi.org/10.1029/92JD00719, 1992.

Rockel, B., Will, A., and Hense, A.: The Regional Climate Model COSMO-CLM (CCLM), Meteorol. Z., 17, 347–348, 2008.

Roeckner, E., Brokopf, R., Esch, M., Giorgetta, M., Hagemann, S., Kornblueh, L., Manzini, E., Schlese, U., and Schulzweida, U.: Sensitivity of Simulated Climate to Horizontal and Vertical Resolution in the ECHAM5 Atmosphere Model, J. Climate, 19, 3771–3791, doi:10.1175/jcli3824.1, URL http://dx.doi.org/10.1175/jcli3824.1, 2006.

Travis, K. R. and Jacob, D. J.: Systematic bias in evaluating chemical transport models with maximum daily 8 h average (MDA8) surface ozone for air quality applications: a case study with GEOS-Chem v9.02, Geoscientific Model Development, 12, 3641–3648, doi:10.5194/gmd-12-3641-2019, URL https://www.geosci-model-dev.net/12/3641/2019/, 2019.