Response to anonymous referee #1:

We thank the referee for their comprehensive and detailed review of our manuscript and provide responses to the questions and suggestions below:

Major comments:

There is too much similarity between the abstract and the conclusions. The conclusion section should benefit from including an outlook paragraph (which, e.g., should not be in the abstract). Now that a study has been done on modifying the chemistry and the resolution, and improvements in the model performance are moderate, what could be the next points of focus for further development of TM5. What would be the way forward to further improve CTMs? Another aspect which could be discussed in the conclusion : as the differences between 1x1 and 2x3 are moderate, is it an option to still use the 3x2 version profiles as a priori for retrievals?

We have now limited our abstract to focus on chemical trace species that will be retrieved by TM5-MP. We modify the abstract accordingly:

We provide a comprehensive description of the high-resolution version of the TM5-MP global Chemistry-Transport Model, which is to be employed for deriving highly resolved vertical profiles of nitrogen dioxide (NO_2), formaldehyde (CH_2O), and sulphur dioxide (SO_2) for use in satellite retrievals from platforms such as the Ozone Monitoring Instrument (OMI) and the Sentinel-5 Precursor, the TROPOspheric Monitoring Instrument (tropOMI). Comparing simulations conducted at horizontal resolutions of 3º x 2º and 1º x 1º reveals differences of ±20% exist in the global seasonal distribution of ^{222}Rn, being larger near specific coastal locations and tropical oceans. For tropospheric ozone (O_3), analysis of the chemical budget terms shows that the impact on globally integrated photolysis rates is rather low, in spite of the higher spatial variability of meteorological data fields from ERA-Interim at 1° x 1°. Surface concentrations of O_3 in high-NO_x regions decrease between 5-10% at 1° x 1° due to a reduction in NO_x recycling terms and an increase in the associated titration term of O_3 by NO. At 1° x 1°, the net global stratosphere-troposphere exchange of O_3 decreases by ~7%, with an associated shift in the hemispheric gradient. By comparing NO, NO_2, HNO_3 and PAN profiles against measurement composites, we show that TM5-MP captures the vertical distribution of NO_x and long-lived NO_x reservoirs at background locations, again with modest changes at 1° x 1°. We show that surface mixing ratios in both NO and NO_2 are generally underestimated in both low and high NO_x scenarios. For Europe, a negative bias exists for [NO] at the surface across the whole domain, with lower biases at 1° x 1° at only ~20% of sites. For NO_2, biases are more variable, with lower (higher) biases at 1° x 1° occurring at ~35% (~20%) of sites, with the remainder showing little change. For CH_2O, the impact of higher resolution on the chemical budget terms is rather modest, with changes less than 5%. The simulated vertical distribution of CH_2O agrees reasonably well with measurements in pristine locations, although column-integrated values are generally underestimated relative to satellite measurements in polluted regions. For SO_2, the performance at 1° x 1° is principally governed by the quality of the emission inventory, with limited improvements in the site specific biases with most showing no significant improvement. For the vertical column, improvements near strong source regions occur which reduce the biases in the integrated column.

Previous studies have quantified retrieval errors with respect to horizontal resolution (e.g. Boersma et al., 2007; Heckel et al., ACP, 2011) and considering the small footprint of the new tropOMI instrument, it seems disingenuous to use a 3° x 2° model grid to perform such retrievals considering the progress made in the instrument resolution.

I have the impression that the manuscript is too strong in its argumentation that emissions are an important reason for discrepancies with observations. I would suggest that the authors make this claim more solid, e.g., by doing sensitivity experiments.

Performing such sensitivity studies would then turn our manuscript into a scientific paper rather than a model description and validation paper, whereas the purpose of our submission to GMD is to provide a peer-reviewed benchmarking reference. It is envisaged that studies related to emission estimates and retrievals from tropOMI will occur once the satellite is launched (spring 2017). Other independent studies have placed discrepancies between models and measurements almost entirely on missing emission terms, therefore allowing inversion studies to be performed (e.g. Elburn et al., ACP, 2007; Kim et al., ACP, 2011; Manning et al., JGR, 2011). Moreover, the basis of emission trend studies from Earth-orbiting satellites relies on the missing component being almost entirely due to emission fluxes (e.g. Schneider et al., 2015). Given the large discrepancy between e.g. lower tropospheric NO_x in Texas, an area subject to high Anthropogenic emissions, the first-order impact is also thought to be from under-estimates in emissions (e.g. de Gouw et al., Env. Sci. Tech., 2011).

The analysis of the differences between the two resolutions should be improved. What about the difference in turbulent vertical transport between both model simulations? What about dry deposition parameterisations affected by the different resolutions? What about lightning NO*x* parameterisations? In the current manuscript, there is a lot of focus on convection, whereas other parameterizations might also play a role.

We focus on the convective aspect as the source of the convective mass-fluxes has changed in this version of the model compared to previous versions, rather than the e.g. turbulent mixing scheme, which is identical. Resolution effects on turbulent mixing would require a separate study and, again, we consider this a model validation paper with a focus on retrievable trace gases, and therefore present the cumulative result of all resolution induced changes. Additional tuning was performed between simulations so that the lightning NO_x is constrained to an annual global total of 6 Tg N yr^{-1} throughout, as described in the text, thus:

For lightning NO_x we use the parameterization which uses convective precipitation fields (Meijer et al., 2001) and constrain the annual global emission term at ~6 Tg N yr^{-1}. This uses the convective flux values meaning that re-scaling of the nudging term was necessary in order to achieve similar total lightning NO_x across simulations.

This ensures that the NO_x emission total is the same between runs allowing a valid comparison.

The vertical grid is identical between 3° x 2° and 1° x 1° simulations. For the dry deposition, although regional terms may exhibit larger differences, the small change in the O_3 deposition term in the Northern Hemisphere given in Table 4 implies this is not a dominating source of the modest differences found.

It is also not made clear how large the differences are in the meteorological fields seen by the different resolutions. E.g., is the total precipitation equal in 3x2 and 1x1? Are the convective mass fluxes, when globally averaged, equal in both versions? Is the cloud cover equal when globally averaged? And the albedo? If not, it would be informative to quantify that.

For details on the use of meteorological fields in TM5-MP the referee is pointed to Bregman et al., ACP, 2003 and Huijnen et al., GMD, 2010. The similarity in both the regional photolysis frequencies (where clouds dominate the total Optical Depth; Figure S4) and the wet and dry deposition fluxes shows that there are no significant changes in the global and zonal mean terms for such quantities. Transport will be better defined using higher resolution wind fields, but this is one of the benefits of increasing horizontal resolution evident in the March INTEX-B comparisons of e.g. O_3. We feel that a comparison of such meteorological fields would detract from the real focus of our paper, which is whether the integrated effect of the change in resolution alters the chemical composition of the troposphere significantly.

As one of the aims of the model is to use it for generating instantaneous a-priori profiles and columns, it is not sufficient in this study to look at biases only. One should also look at the high frequency behaviour and thus, e.g., at correlations. E.g., Table 9 gives seasonal biases, but I think it is necessary to also show correlations. In addition, as the satellite retrievals will be used globally, it is not sufficient to quantify the difference between the 3° x 2° and 1° x 1° versions only for the observation locations of the manuscript. As the (tropospheric) columns of NO_2, SO_2 and CH_2O from TM5 will be important for the retrieval, one could, e.g., estimate how well the 3° x 2° and 1° x 1° distributions are correlated spatially. This could also allow to better quantify whether using 3° x 2° instead of 1° x 1° still makes sense.

As well as presenting the biases at EMEP surface sites in Europe, we also present comparisons of vertical profiles across a wide area from the INTEX-B and Texas-AQSII campaigns (Singh et al., ACP, 2009; Parrish et al., JGR, 2009). The locations chosen for validation are significantly restricted by data availability during 2006. However, the main findings are consistent across all selected regions, therefore we feel confident that as we have compared surface values and vertical profiles in both remote and urban scenarios (i.e. over different chemical regimes) the main biases in any a-priori fields have been sufficiently quantified.

At the request of the referee, we have examined the Pearson correlation co-efficients for the seasonal biases given between observations and instantaneous values at 13:00hrs in Tables 7 and 8. For NO_2, only a few sites exhibit significant correlations with $r > 0.65$ (i.e.) with many more exhibiting anti-correlations i.e. negative r values, especially during DJF, or r values between -0.3-0.3 indicating no meaningful correlation between model and measurements at all. There is typically a marked difference in r between seasons at sites for both simulations, with JJA generally exhibiting higher correlations. Looking across sites reveals increasing resolution does not necessarily increase correlation though, with 1/3 of the sites exhibiting less correlation at 1° x 1°and ¼ being relatively unaffected. Comparing r values at 1° x 1° using the Tiedke convective scheme shows that although there is some impact, there is not a consistent increase in correlation when using the ERA-interim archived mass-fluxes, with many sites exhibiting significant decreases. Therefore, similar to the conclusions regarding seasonal biases, the use of 1° x 1° does not lead to a systematic improvement

in correlation showing the constraints of using monthly mean estimates for emissions towards capturing variability.

We include the following text to summarise this:

Analyzing the corresponding seasonal correlation co-efficients (not shown) shows in ~25% of the cases there is little seasonal correlation between the weekly [NO_2] in TM5-MP and the measurements regardless of resolution for both seasons (Pearson's r in the range -0.3-0.3). In ~30% of cases there is actually a degradation in r between resolutions, the changes somewhat reflect those seen in the seasonal biases i.e. simultaneous changes to both the meteorology and local emission fluxes do not necessarily improve the performance of the model. Comparing 1º x 1º values both with and without the Tiedtke convection scheme shows that for the most convective regions (e.g. south of 45ºN) increases in r generally occur during JJA when employing the ERA-interim mass-fluxes. Conversely for e.g. Finland the correlation becomes worse.

Finally, it is not clear what selection criterium is used for putting some figures and tables in the main document and others in the supplementary material? If a reader decides not to read the supplementary material, he should at least have an idea of what he will miss.

The authors selected which Figures they find most revealing i.e. that show the most interesting findings typical of most manuscripts. We reference the Supplementary Material many times in the text of the manuscript, so assume that the reader has the opportunity to look at all Figures shown if he/she is interested in any particular trace gas.

Specific comments:

page 1, line 15-17 differences ... differences

page 1, line 15-17 increases/decreases : it is not clear which resolution is the reference.

We modify the sentence thus: *Differences of ±20% exist in the global seasonal distribution of ^{222}Rn between simulations conducted at 3° x 2° and 1° x 1°, being larger near specific coastal locations and tropical oceans.*

page 1, line 18 "strength" of convective activity is rather vague. Is CAPE meant, updraft velocity, updraft mass flux?

The archived convective mass-fluxes and detrainment rates are the new meteorological fields employed in TM5-MP from the ECMWF meteorological dataset as described in Sect. 2.1 of the manuscript. We refer to the cumulative changes in convection determined using the ^{222}Rn tracer, which come from a combination of parameters in the meteorological dataset, now summarized as the term "convective transport".

page 1, line 19 NH is not yet defined. What is meant by "NH (tropics)"?

We remove this abbreviation from the abstract and change the text accordingly: *Analyzing vertical profiles of ^{222}Rn above source regions, differences in the strength of the convective transport of between 2 and 10% (~10 and 20%) occur below 700hPa (200hPa) in the Northern Hemisphere around the tropics.*

page 1, line 20-21 from simulations at 1x1 horizontal resolution. Isn't it also done for 3x2?

To determine any difference in J values requires the comparison of two different runs. We clarify this in the abstract thus: *For tropospheric ozone (O_3) analysis of the chemical budget terms between simulations shows that the impact on globally integrated photolysis rates is rather low, in spite of the higher spatial variability of meteorological data fields from ERA-Interim at 1° x 1°.*

page 1, line 31-32 not clear whether for both resolutions.

We change the text accordingly: "*By comparing NO, NO_2, HNO_3 and PAN profiles from both simulations against a host of measurements … "*

page 1, line 34 shouldn't 20 and 35 sum up to 100? At this stage, the reader is not yet aware of the fact that changes of less then 5% are not accounted for.

We imply that at 45% of the sites there is no significant change in the bias. We change the text accordingly: *For NO_2, biases are more variable, with lower (higher) biases at 1° x 1° occurring at ~35% (~20%) of sites, with the remainder showing little change.*

page 1, line 35 in TM5-PP : only the high resolution version.

Figure 8 shows that there is a seasonal cycle in [HNO_3] for both simulations, where there is a strong correlation for between simulations.

page 2, line 55 The text mentions current resolutions of 2–4◦ in latitude and 2–6◦ in longitude. There are however currently models with higher resolutions, see, e.g., Yu et al. [2013.]

We thank the referee for this information and update the text accordingly.

page 3, line 113 which replaces the parameterization of Tiedtke (1989). Be clearer about what sub-grid scale parameterizations are still calculated in TM5. E.g., are turbulent diffusion coefficients calculated, or have they been archived too?

We now include the following text: "*The vertical diffusion in the free troposphere is calculated according to Louis (1979), and in the BL by the approach of Holtslag and Boville (1993). Diurnal variability in the BL height is determined using the parameterization of Vogelezang and Holtslag (1996)."*

page 3, line 117-119 Concerning large scale transport, one mentions the CFL criterium. In addition, maybe it is interesting for the reader to mention which transport scheme is used.

We change the text accordingly: *"We use the first-order moments scheme with an iterative time-step to prevent too much mass being transported out of any particular grid-cell…"*

page 6, line 231 The Schery et al. (2004) reference for the 222Rn emissions is difficult to find, as it is part of a book. It would therefore be useful to describe shortly some aspects of the 222Rn emission map: is it only from continents, is there a latitudinal gradient, are there emissions at high latitudes, is it very patchy or rather homogeneous? Is it dependent on soil moisture or precipitation?

The distribution of global ^{222}Rn emissions is shown in Zhang et al, ACP, 2011. We now reference this publication for readers interested in the specifics of the ^{222}Rn distribution. We feel that an in-depth discussion of the emission inventories used detracts from the main focus of our manuscript.

page 8, line 290-293 Differences are attributed to the resolution of the emission data set, and the convection. It is not clear why the temporal resolution of the emissions should play a large role. Isn't it mainly the horizontal resolution which plays a role in explaining the difference between 3°x2° and 1°x1°? In addition is written earlier in the text (page 8, line 283) that 222Rn is emitted at a steady rate.

The emissions, which are typically provided at 0.5-1° resolution, are distributed onto the working model grid. Therefore more heterogeneity occurs on a higher resolution as urban and rural centers are differentiated more acutely. Emission at a steady rate means there is no variability in the monthly mean emission flux representing meteorological factor or diurnal variability.

page 8, line 301-303 are the globally averaged 222Rn emissions equal in 3° x 2° and 1° x 1°? Are the globally average convective mass fluxes equal in 3° x 2° and 1° x 1°?

page 8, line 304-306 I would expect that, if archived mass fluxes are used, the global total mass flux is equal, independent of the resolution (1° x 1° or 3° x 2°). Secondly, if mass fluxes were stronger in 1° x 1°, I would expect for 1° x 1° (compared to 3° x 2°) lower ^{222}Rn concentrations at the surface and higher concentrations between 900 and 700 hPa. But for DJF in Paris and London one sees the inverse

All globally integrated emission fluxes of ^{222}Rn are identical between simulations allowing a valid comparison of results, similar to the other emissions introduced into TM5-MP. The values at specific locations do change though due to the degree of coarsening of the 0.5° x 0.5° ECMWF data needed for the different resolutions (although the area-weighted total is equal to the original ECMWF data in both cases). For comprehensive details on the use of meteorological datasets in TM5 the referee is referred to Huijnen et al. (2010), which for the sake of brevity we do not include in our manuscript. As would be expected, the global mean of the convective mass fluxes calculated using 3° x 2° and 1° x 1° values can be slightly different due to potentially wider spread in the 1° x 1° values (more members of the data array), although the summed total will be equal. This holds for other tropospheric parameters such as temperature and surface albedo. Here we are more interested in regional differences. To remove the variability in emission fluxes above point locations averaging of the 1° x 1° profiles is necessary (where decomposition of the 3° x 2° profile at sub-grid scale is not possible), thus being able to differentiate the impact of the meteorology. Under instances of weak convective activity (DJF), our results show that indeed the coarsened 3 x 2 convective mass-flux can result in more uplift than at 1° x 1°, due to the variability in the averaged 1° x 1° values being high.

page 9, line 341 and around Iceland for JJA. This difference is hard to distinguish.

We remove this from the text and will provide a higher resolution version of the diagram to improve clarity.

page 9, line 344-346 Here no averaging is performed towards an identical hor … : does it mean that the value of a 1° x 1° grid box is compared with the value of a 3° x 2° grid box? Is there a spatial interpolation between 3° x 2° grid points, and 1° x 1° grid points?

We use the geographical location of the cities to perform interpolation in both cases, as for all the profile comparisons for trace gases shown in the manuscript.

page 10, line 368-369 Is this relevant as only 90◦S-30◦S, 30◦S-30◦N, and 30◦N-90◦N are shown?

Yes, because these three zones are comprised of cumulative sums from the 10° bands, therefore we inform the reader as to the resolution of the budget terms.

page 10, line 372-373 The abbreviations O3S (tagged O3 tracer which undergoes only …) and BO3S (stratospheric burden of O_3) are confusing. If O3S is a tracer as defined above, it would be logical that BO3S would be just the total burden of that tracer, whereas here that is not the case.

BO3S is the burden of O3S, thus : *"The stratospheric burden of O_3 (BO_3S) exhibits a strong hemispheric gradient … "*

page 10, line 390-391 "At 1°x1° the largest increase in STE occurs in the SH during JJA." : This cannot be seen from the numbers in the tables. Isn't the aim explaining the 7% reduction?

Figure 2 is introduced at the start of this paragraph (line 387) and we refer to this when discussing the change in the latitudinal gradient in Stratospheric O_3 with respect to the downwelling.

page 11, line 414-417 Differences are attributed to only 4 causes. Might there not be an impact of the resolution on the dry deposition, the turbulent mixing, the large scale transport, or the mass conservation of the transport?

The fact that the cumulative deposition velocities for e.g. O_3 (Table 3) are essentially the same between the 3° x 2° and 1° x 1° simulations shows that dry deposition effects are minimal. We implicitly examine the differences in the turbulent mixing with the Rn^{222} comparisons, which hold for O_3 considering the tropospheric lifetime is typically > 20 days. We modify the text accordingly. The large scale transport does change as shown in the INTEXB comparison and we comment on it there. Without performing tagged O_3 experiments we cannot fully quantify changes in the long-range transport component.

page 12, line 428 shows that differences are small : across resolutions?

Between the different simulations thus resolutions. We change the text thus:" … *shows that differences are small between simulations, and typically mimic those which occur at the surface."*

page 11-12, line 436-438 this does not seem to hold on a regional scale.

All budget terms we show are for the global or a zonal domain. There is no 3D budget file output during a run due to computational constraints. We realise that providing analyzing results in this way will not provide exact changes in clean/polluted regions. However, the comparison of O_3 mixing ratios in Europe again EMEP measurements shows that differences between resolutions are small,

therefore changes are no so large as to lead to first-order reductions in resident [O_3] due to much higher [NO].

page 13, l 501 What is ORGNTR?

This is the tracer name for lumped alkyl nitrates. We now include a definition at the end of the introduction along with HNO_3 and PAN.

page 13, line 482-483 In the lower troposphere (<900 hPa) : as it is written here, one interpretes it as if (<900 hPa) is the definition of the lower troposphere. It is however meant to be an extra condition.

Rather than referring to a designated definition, we only use the terms to describe our conclusions on what is shown in Fig.6.

page 13, line 490-493 And what about October?

This is now corrected.

page 13, line 504 to quantify the effect on higher spatial resolution.

Now changed.

page 13, line 515 decrease marginally by 2-3% : looking at Table S1, shouldn't it be 1-3%?

Now corrected.

page 15, line 581 splitting the atmosphere in 3 regions (NH extra-tropics, tropics, SH extra-tropics) is much rougher than "zonally integrated"

Zonally integrated refers to the cumulative values across all longitudes.

page 16, line 603 Fig. S13 : is this the correct figure to be referred to?

We have now corrected to text referencing the correct Figure.

page 16, line 616-617 Isn't it the increase in spatial resolution which helps?

Any improvement in the temporal distribution comes from an using a higher horizontal resolution on which gridded emission estimates are applied. We now change the text to: " *At 1º x 1º significant improvements occur as a result of the better temporal resolution of the emission sources as a result of increasing horizontal resolution.*"

page 16, line 619 aggregated on a weekly basis does not matter if one looks at seasonal biases; it would have played a role if one also shows correlations.

The value presented is a mean seasonal bias as derived using the bias values from weekly points rather than a single seasonal value.

page 24, Table 4 Whereas most terms in this table are in units of TgO3 yr−1, it is unclear in what units BO3 and Strat. BO3 are. If these are burdens, one would expect Strat. BO3 to be a larger fraction of BO3, than the values shown here (e.g., on the global scale 80 and 378).

We now include in the Table heading : " *.. with all quantities being given in Tg* O_3 yr^{-1}." We take our number from the individual budget files in order to quantify our Strat. BO3. Figure 2 shows that the zonal mean ratio is between 0.05-0.7 in the troposphere, with the higher ratio correlating with lower air pressure thus less mass. Given that through most of the troposphere the ratio changes between 0.05-0.3, 80 Tg seems a reasonable total.

page 24, Table 4 It is not clear where one can find "The fraction of the tropospheric burden originating from the stratosphere is also given." Does one mean Strat BO3? Are these absolute values Tg, or is it %?

We now remove this from the Table legend.

page 24, Table 4 Why no % for the NH/SH/tropical STE changes? As these values are not given, the sentence on page 10, line 379 is rather unclear : "The increase in STE in the SH, with an associated decrease in the NH ..."

We now add the percentage differences for each chosen zone.

page 25, line 744-745 Those with differences <5% are considered to exhibit no discernible change in the bias. An interpretation should not be written in the caption of a figure.

The final print version of this Table will include coloring such that the number of positive and negative biases >5% can be discerned quickly. The policy of GMD is not to include colour in the text in the first instance. The <5% comment relates to the fact that the (black) entries in the table represent stations that are essentially unchanged. Therefore, rather than a definition is pertains to entries in the Table.

page 31-32, Figs. 4-5 Are the observations also just the 13:00 values?

Yes, please see Sect. 2.3.

page 35, line 823 during September 2006 : but October 2006 is apparently also shown.

Figure legend now corrected.