Review #1

General reply to reviewers

We thank the reviewer for helpful comments that make the paper clearer. Below we answer general and specific comments, but first we address some necessary manuscript changes due to correction of model/diagnostic errors identified during the course of responding to the reviews.

Important manuscript changes

1. During the review process, we located and corrected two errors in the CTM3 nitrate aerosol module. First, wet scavenging of these aerosols was too slow and when correcting this less HNO3 is removed through wet scavenging of nitrate aerosol. Second, nitrate aerosols were wrongly affecting stratospheric NOy/NOx, in effect reducing reactive NOx in the UTLS. Correcting this lead to about 10% higher tropospheric O3 burden in CTM3 and also affects the OH and hence CH4 lifetime. STE now does not change when including nitrate aerosols. When corrected, the effect of including nitrate aerosols in CTM3 is similar to that in CTM2. The discussions of OH, CH4 lifetime, O3 burden and STE are revised accordingly, and statements about nitrate being more effective in CTM3 have therefore been removed. These corrections do not change the conclusions of the work, but makes the results more consistent.

2. We have removed the lightning sensitivity section 4.2 and condensed the text into the lightning section 2.6. Figure 11 is therefore removed, along with the entry C3_PIC in Table 2. One additional simulation has been carried out with a fixed vertical scaling of L-NOx approximately similar to CTM2, revealing this to account for about 50% of the CTM2-CTM3 differences in OH, CH4 lifetime and O3 burden.
   The new lightning paragraph reads:
   "We evaluate the new L-NOx emission parameterization against a simulation that uses the previous vertical emission profiles. These old vertical profiles (Pickering et al., 1998) injected a large fraction of L-NOx near the surface and near the convective cloud top. In contrast, the new vertical profiles (Ott et al., 2010), which are based on more extensive in situ measurements, place most L-NOx below the convective cloud top with little near the surface. While the new L-NOx algorithm scales the new profiles to match the cloud-top heights hourly in each CTM column, the old CTM2 algorithm assumed a fixed convective top at 16 km, for purposes of calculating the vertical distribution of L-NOx. Thus CTM2 placed much more L-NOx in near the tropopause, or even in the stratosphere when the tropopause was below 16 km. The change in L-NOx between CTM3 and CTM2 is extensive, including location, flash-rate and scaling factors as well as vertical profiles of the injected NO.
   We have tested the change in profiles in CTM3, and cound that compared to the old profiles (Pickering et al., 1998), the new Ott et al. (2010) profiles cause modelled zonal mean NOx to increase by 10-15% (annually up to ~10%) in the middle troposphere (400-800 hPa) and decrease NOx by up to 15-25% (~15% annually) in the tropical upper troposphere (250-100 hPa). Accompanying O3 changes range from -2% to 2%. However, scaling the old profiles to fixed 16 km convective cloud tops has a larger effect, increasing NOx} by more than 100 % at ~ 200 hPa and O3 by ~35 % at 400 hPa. We come back to this in Sect. 3.3."

3. Because the lightning was removed from section 4.2, the transport sensitivity section 4.1 is now called section 4, "Transport sensitivity studies". It has been updated to take into account reviewers comments.

4. Simulation names are changed for clarity:
   C3_30MIN -> C3_1/2 (halving the operator split time step, i.e. 30min)
   C3_V2 -> C3_pole
   _NIT -> _ssn ('ssn' to reflect sea salt, sulfate and nitrate; less focus on nitrate)
   In addition we have included C3_1/4, where the original time step is quartered (i.e. 15min).

5. Table 2, showing age of air runs, has been removed, because the information is already in Table 1. Where needed, the text has been changed to reflect this.

6. We have shortened the CH4 lifetime section, to make it clearer. And we have changed the focus of O3 burden to the 150ppb O3 definition, to better be able to compare with Stevenson et al.

7. Wet scavenging section: The two paragraphs before the last, are modified to match the correction of nitrate aerosols:
   "Our chemistry simulations C2 and C3 show similar results; lessHNO3 is scavenged in CTM3, giving the latter a higher tropospheric burden of HNO3. The very different pattern of scavenging from CTM2 to CTM3 also contributes to this, by changing the relative distribution of HNO3 throughout the troposphere. In the ssn-simulations, we see a reduction of tropospheric HNO3 by about 28% in CTM2 and 32 in CTM3. This agrees well with what we found for OH and lifetime ofCH4. The slightly higher effect on HNO3 in CTM3 is probably due to the less efficient wet scavenging in CTM3 also affecting sea salt and nitrate aerosols, allowing more HNO3 to be bound in nitrate particles. Also important are differences in other loss processes such as photodissociation and the loss to OH.

   The higher HNO3 burden in CTM3 is not distributed evenly. In the tropics, below 400hPa, HNO3 is about 50% lower in CTM3, while in the extra-tropics CTM3 has up to twice as much HNO3, consistent with less scavenging. Inclusion of tropospheric aerosols reduces extra-tropical HNO3 by up to 100 % in both CTM2 and CTM3, slightly more pronounced in CTM3. Because of the less efficient scavenging of sea salt aerosols in CTM3, more HNO3 is taken up to form nitrate aerosols, explaining this difference. Above 200 hPa, in the stratosphere, CTM3 has about half of the HNO3 in CTM2, except in the upper stratosphere, where CTM3 again has somewhat higher HNO3."

Review #1

General comments:
I would like to see the authors distinguish more clearly between the model and the concepts on which it is built from the ECMWF winds that are input to the model. The computer construct could be perfectly adequate but some deficiencies (like STE) cannot be remedied except by improving the input to the model.

This is a good point. Certainly the conceptual model that we encoded as CTM3 should have all the processes and numeric necessary to calculate the STE O3 flux, but if the met data from the ECMWF forecast has noise then the STE will be incorrect. Similarly, we have updated the scavenging to follow the precipitation and cloud fraction, but if the physical-chemical processes we are using (from lab and field studies) do not adequately describe the real processes, then the CTM concept is fine but fails due to lack of chemical data. In general it is not always easy to separate these errors – that of failed concepts from that of incorrect input data.

There are numerous grammatical errors, some of which I point out, and I encourage one of the native English speakers to read the final version with this in mind.

We thank the reviewer for localizing these errors, and we have read carefully through the paper to double check for other grammatical errors.

Resolution – The discussion of spatial and temporal resolution could be much more clear. The long time step without an increase in transport error seems to contradict statements about errors near the jets and the polar vortex. When high resolution simulations are discussed, I think this means high spatial resolution (because the longer time step allows it) but this needs to be clarified. I was confused by the prior statements about transport near the jets and the possibility to reduce the time step if needed when reading the description of advection and the internal computation of the time step.

We have changed the text and added some examples of error and time step to make this clearer. The new tracer-transport structure in CTM3 reduces error in vertical transport by combining w and convection; it allows for CFL violation in the horizontal advection. Basically, the CTM3 is now stable for long time steps compared

with CTM2 (this is very important, since instability is a worse problem than numerical error). The long time steps induce larger numerical errors, but these converge quickly as the time step is halved. With long time steps (60 min) the transport errors occur only where the jet turns, i.e., where the air parcels in the jet are moved beyond the turning point and overshoot the jet.
The advantage of the new transport is that the global time step is not limited by the grid-cell divided by the jet speed as it is in CTM2. Thus CTM3 can afford to be run at 1° or finer with a 15-30 min time step, but it will have larger transport errors where the jet turns. By allowing ready calculation at different time steps, however, CTM3 allows error estimation and correction unlike CTM2 and most all current CTMs.

The paper should summarize the reason for using pieced forecasts rather than the analysis fields and how this affects meteorology. There are clearly some questions about the meteorology and deficiencies in the current system compared to the previous system that need to be discussed more directly.

We have included the original CTM references that discussed this, and only briefly note here that:
"It has been clear in the history of ECMWF forecasts that there is a spin-up time for the forecasts during which short-lived phenomena like precipitation adjust to inadequate initialization from the analysis fields (Wild et al 2001, Kraabøl et al, 2002). Forecasts also allows us to retrieve additional meteorological data such as convective mass fluxes, which are not available from operational analyses. The 12-hour spin-up used here was based on ECMWF experience. The use of pieced forecasts is also imperative when running an externally forced CTM because much of the physics need to be integrated, not snapshots, (e.g., 3-D clouds, convective fluxes, and precipitation fields), which are integrated from the forecast model. If analysis fields are used as the core meteorology in the CTM, then the CTM must include the full general circulation model physics (unlike here)."

Forecasts are used for two main reasons; the meteorological fields are more consistent (Wild et al 2001, Kraabøl et al (2002)), and we require additional output from the ECMWF, namely convective mass fluxes. In addition we produce these data each 3 hours.
New text:
"Forecasts allows for a more dynamically consistent dataset (Kraabol et al, 2002; Wild et al., 2003), and also allows us to retrieve additional meteorological data such as convective mass fluxes, which are not available from operational analyses."

When comparing the versions of the CTM it isn't always clear whether meteorology is the same or different.

It is always the same, and we have included this more specifically.

Do any observations compare sufficiently differently with the simulations that we can unequivocally tell the simulations apart (and pick one as better?). Even in the abstract – compares with observations 'well'. There are many improvements – is the new CTM objectively better or similar but faster?

We have changed revised these statements, e.g. in the abstract:
"Compared with Oslo CTM2, Oslo CTM3 is faster, more capable and has better conceptual models for scavenging, vertical transport and fractional cloud cover, but shows no significant improvements in terms of matching atmospheric observations."

The figures are often small and difficult to read even if enlarged. Perhaps they can be made at higher resolution so that enlargement makes them more legible.

We have improved the figures.

Specific Comments:
1562 Abstract Line 6 – high spatial resolution
Corrected

Line 10 Use of the same meteorology to drive the two models shows that
Corrected

1563 Line 15 first time the resolution issue. Even in the abstract – you are losing spatial information in regions of sharp gradients due to the long time step – so what do you gain from more spatial resolution? Need to be clear.

We have clarified this paragraph with respect to resolution (fast vs. accurate), and moved it to the conclusions, as we agree with reviewer #2 that it is misplaced.

1565 L 13 'stratospheric losses' – I presume this means long-lived gases, and the radicals and reservoirs aren't considered at all? (which is fine for this application, but since the purpose of this paper is to describe the model I think the statement should be explicit).

It applies to hydrocarbons and CO. When stratospheric chemistry is turned off, the most long-lived species is CH_4. We have added this information to the text.
"In such cases, all the tropospheric trace species are advected throughout the stratosphere but without real chemistry: species that are photochemically destroyed in the stratosphere are allowed to decay at a fixed rate; where species with sources in the stratosphere like O_3 and NO_x are set to climatological values at CTM levels a few km above the tropopause."

L 14 – I don't understand why the statement about the CTM2 climatology is here at all.

We agree and have removed the sentence.

1567 L 15ff – This paragraph is confusing. The P2008 connection was severed? By the meteorology or by the implementation? The new meteorology is good or the new transport is good. The pie shape boxes have a problem in V2 (which is the new algorithm) – what is the difference between the new algorithm (V2) and the 'final algorithm'. Is the final algorithm an option or the new standard algorithm?

The meteorology has not changed from CTM2 to CTM3, only the transport code. P2008 combines the opposite polar pie grid boxes to reduce the time step necessary and more importantly to allow for over-the-pole flow that is important in solid-body rotation tests (see P2008). CTM3 originally took the P2008 tracer algorithm as published, but the examples quoted above showed that the polar transport had some undiscovered flaws. Thus several tests with the P2008 SOM code led to an improved treatment of polar advection that more accurately reproduced both cyclonic and anti-cyclonic vortexes with edges over the pole. We have revised the text of this paragraph to be clearer.
New text:
"In the updated, improved SOM version of the P2008 advection algorithm used in CTM3, meridional advection is still calculated as a connected pipe-flow combining a meridion with its antipode complement, but over-the-pole flow at the two pie-shaped grid boxes at each pole is no longer implied. These two grid boxes are no-longer combined before the [V] advection step, and all transport in these polar boxes is through the [U]-flux. Given the very small area of these polar-pie boxes, the Lifshitz limiter severely restricted the size of the global time step. Thus each polar-pie box and its adjacent lower-latitude box are combined (conserving all moments) before both [U] and [V] transport, and restored to individual boxes after. An alternative version of the SOM polar advection, in which the polar-pie boxes are not combined with their much larger lower-latitude neighbors,

is tested here and designated C3_pole. In the optional C3_pole version, the Lifshitz limiter forces global time steps of 30 min, instead of 60 min, and thus the polar transport is more accurate because of higher spatial resolution at the poles and shorter time steps. ”

The following paragraph (“The V2 transport….”) was confusing and unnecessary and was dropped.

1569 L 8 the amount of tracer solved in rain? Technically correct I think (I looked it up) but not the most common usage (I checked).

Yes, we have fixed to:
“In general the ratio of tracer dissolved in rain to that in interstitial air is calculated using Henry’s Law.”

1570 L 17 – what ‘new’ solar flux data are you using?

The revised solar fluxes are now an average of SUSIM high and low solar activity measurements as described in Hsu and Prather (2009, JGR). That reference describes the updates to the photolysis code and is now inserted:
L19: “Cross sections, quantum yields, and solar fluxes are as described by Hsu and Prather (2009).”

L 24 there is a problem with the sentence that starts ‘With the JPL-2010 update – should ‘Sander et al.’ be in parentheses and follow ‘and more complex wavelength-pressure dependence”?

Yes, the referencing is confusing:
L23: “With the JPL-2010 update (Sander et al, 2011) most VOCs are now included, and the next release of fast-JX…”

1575 – when talking about the lightning – ‘the high flash rates are better modeled’ – I am not sure where the ‘better’ judgment comes from. Physical basis of parameterization or some other comparisons with observations?

We have rewritten the comparison between L-NOx in CTM2 and CTM3 to address the reviewers concern. The text now says:

“Figure 1 compares the lightning distributions in CTM2 and CTM3 with OTD-LIS observations. CTM3 has much less lightning over oceans than CTM2, especially around Indonesia, but still more than the observations. In both models, flash rates are underestimated in Africa and overestimated in South America, but these biases are slightly reduced in CTM3. CTM3 also better reproduces the seasonal north-south migration of lightning in Africa, as seen in the monthly standard deviations (Fig. 1, bottom row). Overall, the L-NOx spatial distributions in CTM2 and CTM3 are similar, but CTM3 better matches the ODT-LIS observations.”

In addition we have rewritten some paragraphs in this section to make the text shorter and clearer.

1576 age of air – Are CTM3 and CTM2 run with the same meteorology? It seems so but an explicit statement would be useful.

Yes, they run with the same meteorology. We have clarified this in the text.

1577 – The age discussion – no data brought to bear anywhere, so I guess changing the transport timestep makes the simulations 'agree' but it isn't clear to me that anything is 'better'.

1576 L20- revised to clarify:
"In the tests here with CTM2 and CTM3, the purpose of examining age of air is to identify differences in the tracer transport algorithm and the impact of polar treatment and time step. All versions use the identical three-hour averaged meteorological data. Our intent is not to evaluate these met fields against observations of tracers that approximate the age of air (e.g., SF_6, CO_2), but to examine how different numerics can produce different values (e.g., P2008). We calculate stratospheric age of air from a tracer that is forced to be linearly increasing in the lower tropical troposphere (Hall, 1999). The age….."

1577 L 21-24 – this paragraph was confusing and not relevant to age-of-air, so it was dropped

CCMVal (and other papers) show a more quantitative comparison for polar ozone loss. In general – ''reasonable polar ozone loss' – statements like these should be avoided in a model description when quantitative metrics are available.
Whey you say 'better' than Sovde et al. (2008) is this subjective or objective?

The 60-layer meteorological data produce a better defined mini-hole than the 40-layer data used in Søvde et al. (2008). The improvement can clearly be seen in north-east of Europe, towards Siberia. We have included a statement about this:
"Both CTM3 and CTM2 reproduce this event well and to a better degree than in Søvde et al. (2008), which had too high O_3 columns in the northern Europe towards Siberia. This improvement is mainly due better stratospheric circulation in the 60-layer meteorological data used here compared to the 40-layer data used in Søvde et al. (2008)."

1579 – Is the HNO_3 problem a lack of NOx or is it a lack of processes to convert NOx to HNO_3? (there is a paper by Kawa et al. discussing this in UARS data).

In the CTMs we include conversion on aerosols and PSCs, but not on ion clusters. Several papers have studied the latter and found it to be the cause of the upper-stratosphere maximum of HNO_3. However, enhanced NOx is needed to form N_2O_5 to be processed on the ion clusters. This NOx can be produced by solar proton events, but some NOx may also be transported downwards from the mesosphere. The CTM will not catch the latter due to the lack of mesospheric chemistry. We have added some more discussion on this, to make the text clearer.

1579 – stagnant vertical transport – what about previous discussion of age of air?

As corrected earlier, the age of air comparisons are model-model version-level comparisons using identical met fields. Age of air cannot directly diagnose the "photochemical" age (i.e., N_2O) since the former includes the long-lived tail for which the N_2O is long gone. We think the N_2O comparison here with MLS is a clear diagnostic.

1580 – N_2O gradients matter at least as much as the N_2O magnitude in developing the FRIAC. I expected a better explanation of why these results are not as good as discussed by Allen et al. for a CTM that is conceptually similar. It is likely that the 'stagnant vertical transport' matters to this also.

We do not agree that the results are not as good as in Allen et al. Allen et al. stop the color scale at 100ppb, which affects what the eye sees as good agreement. When this is taken into account (our scale stops at 175ppb), CTM3 produces similar values as their Replay, well into July, but not as high in August.
The model bias in CTM3 (from slower, stagnant vertical uplift in the tropics in the ECMWF met fields (both CTM2 and CTM3)) means that the absolute values of N_2O are shifted downward. The FRIAC at these altitudes

occurs on times scales of a few months and thus is for the most part determined by the initial, frozen-in N_2O distribution in March. Chemical evolution and the BDC have smaller impact.

We have added contour lines for 75ppb and 100ppb for the CTM3 and CTM2 figures, showing this more clearly, and also improved the text to make the comparison clearer.

1584 Line 15 – drop ‘carried out’ (confuses the sentence and is not needed).

Done

1585 CO emissions are too high, too low – how much OH difference can you account for by changing the emissions (within reason).

We have carried out a 14-month test with 20% higher anthropogenic CO emissions, resulting in less than 1% change in the global OH number. Text is now corrected and this new run is noted.

In other words, a possible low CO from anthropogenic emissions is probably not responsible for high OH.
We have added some text on this in the manuscript.

1586-87 The discussion of the OH lifetime could be much more clear. The ‘literature’ values are based on what? How does Prather get his estimate, and why is it different from the literature. It seems to me you need to say something about that if you are going to use these values as some marker for the difference in simulations. In some sense – literature and Prather differ by more than C3 and C2 – so is the difference significant at all?

Text is correct as is, the papers are all published, but text is revised below to clarify:
“Recent evaluation of the CH_4 lifetime (Prather et al. 2012), total and against loss by tropospheric OH, take into account the recently observed CH_3CCl_3 decay (Montzka et al., 2011) and up-to-date estimates with uncertainties of the other factors that go into the lifetime calculation.”

Montzka, S. A., Krol, M., Dlugokencky, E., Hall, B., Jockel, P., and Lelieveld, J.: Small Interannual Variability of Global Atmospheric Hydroxyl, Science, 331, 67–69, doi: 10.1126/science.1197640, 2011.

Prather, M. J., Holmes, C. D., and Hsu, J.: Reactive greenhouse gas scenarios: systematic exploration of uncertainties and the role of atmospheric chemistry, Geophys. Res. Lett., 39, L09803, doi:10.1029/2012GL051440, 2012.

1587 if stagnant meteorology makes the N_2O lifetime long, won’t it have the same effect on the stratospheric part of the CH_4 lifetime?

Yes, but the stratospheric part of the CH_4 lifetime is only a small fraction of the overall lifetime. From the table, one can estimate that the CH_4 lifetime against stratospheric loss in CTM3 is about 150 yr, which is longer than the canonically agreed upon lifetime of 120 yr (see Prather et al 2012). This offset is the same as for N_2O, but has little impact on the overall CH_4 lifetime, which is dominated by tropospheric loss.

1589 500 Tg O3/yr from Hsu et al. 2005 (among others). Is this the same method used here? The values here are much smaller. The big increase when removing PSCs is odd – the observational estimates don't make such a distinction – does the Hsu et al. estimate include PSCs? This is a confusing discussion with no conclusion.

Yes, we use the same method as in Hsu et al. 2005, as described in the text. There are two main reasons for the lower values compared to their study: We calculate stratospheric chemistry, and the meteorological cycle is different. The effect of the meteorological cycle is discussed, and at UCI found to be giving lower STE than the older cycle.
The effect of using a stratospheric chemistry scheme versus Linoz (as in UCI-CTM) has not been tested.

Hsu and Prather (JGR, 114, D06102, doi:10.1029/2008JD010942, 2009) found by using Linoz with and without PSCs could reduce the STE from 613Tg/yr to 515Tg/yr, i.e. a reduction of 97Tg/yr. In terms of burden, we get similar reduction.

Using the 2005 methodology, the ozone hole complicates the diagnosis because the 120 ppb surface goes far into the stratospheric polar vortex, confusing what is stratosphere and what is troposphere.

We have revised the PSC discussion: Also See point below the next.

1589 – if the meteorology has a 'stagnant Brewer Dobson circulation' no change in the ozone isopleths is going to give a better STE estimate – or a better age of air or appropriate mass flux. If the upward transport is slow, the downward transport (of mass and ozone) is also slow. This isn't speaking about the model – C2 or C3, it doesn't matter. There are issues with the met fields that make them inappropriate for many studies – probably for example included the FrIAC, because no matter the numeric if the gradients are not appropriate in the middle stratosphere to begin with, the initial pulse of tracer that forms the FrIAC will have insufficient contrast with respect to the background and compare poorly with observations. A few well crafted and specific sentences on this point are needed – and along with this fewer general statements.

We have changed "stagnant" with the more appropriate "slower Brewer-Dobson circulation". This we see in N2O vertical profiles, but also STE is affected; SH STE drops to 100, NH drops to 225 in UCI (Hsu and Prather, 2009). Changing the O3 isopleths will change the diagnosed number, but we agree that it does not necessarily give a better STE estimate.
We have revised the text to be more specific. (See answer above and next comment)

1589 – why would it make sense to take out the polar chemistry when calculating STE? The STE of ozone in the true atmosphere has an ozone hole (and dilution etc.) affecting ozone isopleths throughout the SH in particular.

We agree that this paragraph is confusing. Most models contributing to the Stevenson STE comparison do not include comprehensive stratospheric chemistry, and will therefore not be able to accurately model the stratospheric NOx and O3 due to PSCs. Our models may therefore produce lower O3 in the lower stratosphere and thus a smaller STE.
It may also be that STE based on measurements does not fully take the PSCs into account, or at least that it is not representative to compare it with model results. Ideally, models should infer flux in the same way as for observations. We have made the text clearer to discuss this.

New text:
"In an atmosphere without polar stratospheric clouds (PSCs), the 120 ppb isopleth can be very different than when PSCs are included, and stratopsheric O3 is higher. Hsu and Prather (2009) found PSCs to reduce STE by 97 Tg(O3)/yr, and when testing this in CTM3 we find the STE to increase to 427 Tg(O3)/yr for the year 2005. It may be that STE inferred from observations does not fully take the PSC-O3-hole into account, so modelled STE without PSCs provide insight in this. However, it should be kept in mind that the 120 ppb isopleth has changed."

1597 - the new model is as good as if not better than the old model – you have made changes to improve implementation (e.g., numerical transport) and to improve the physical basis of some parameterizations (e.g., lightning and convection). Even if the comparisons with observations were nearly identical, you could argue the new version is 'better' than the old version because the physical basis is improved, at least in the troposphere (which you finally do in the last two sentences of the paper). There is some confusion (in my opinion) about the stratosphere because of the new version of the met fields – that 'stagnant' meteorology depends on input and not on the model per se. However, some aspects of the comparisons with data are affected.

As already mentioned, the met fields are not new, CTM2 and CTM3 are driven by the same meteorological data, and the slower Brewer Dobson circulation, particularly in the tropics, is found in both (we have changed "stagnant" with "slower BDC").
See answer to general comment – the new CTM3 is more capable, and more realistic in some of the physics.

1597 – Discussion the OH – Present model should be improved, but OH is high. Then representing other processes (sulfates and nitrates), the 'bias' is reduced. There could be somewhat more cogent discussion on this point – the version with less physical representation of precipitation scavenging had 'apparently' better OH but clearly if the sulfates and nitrates were not included the good agreement would be for the wrong reason.

Thanks for pointing this out, we have included some more discussion on this point:
"Including sulfur chemistry and nitrate aerosols increases the CH4-lifetime by ~12% to 8.9 yr. We find that increasing anthropogenic CO emissions in the RETRO dataset by 20% cannot reduce OH in CTM3 by more than 1%, and does not affect the CH4-lifetime much. With the improvement to fast-JX and the new scavenging in CTM3, it may be that CTM2 has a longer CH4-lifetime for the wrong reasons, e.g. by distributing lightning NOx to too high altitudes."

1598 – it is a serious point to discuss how to fix up the STE without fixing up the circulation – in the spirit of the preceding comment (right answer, wrong reason) I would say that it is not possible to have the correct fluxes of mass and ozone without having a realistic stratospheric circulation. Those other things that could increase the amount calculated for STE are secondary to this point.

We agree; we cannot fix STE without redoing entire ECMWF IFS model. Likewise, we cannot fix too slow uplift in tropical stratosphere. We are in the process of producing meteorological data from a newer ECMWF cycle, which will shed more light into this.

Figures
Figure 4 – if the point is pattern recognition – fine, although masking the output so it isn't shown where there are no observations would make pattern recognition easier.
Nonetheless – differences between TOMS and both CTMs are obvious, and it isn't clear one is better than the other.

We have made it clearer in the text that no conclusion can be drawn on whether CTM3 is better than CTM2. The purpose of the figure is mainly pattern recognition, to show that main features in the column are reproduced by the models. From a modeller's perspective, we find it important to include the regions where TOMS does not provide data, but we have included a line in the figure to show this area in the model data.

We have included a new figure on CTM3-CTM2 difference for different operator-split time steps, showing how CTM3 converges towards CTM2. We also note that CTM2 does not necessary provide the truth.

Figure 5 – the feature you are discussing (the mini hole over Europe) is nearly invisible in this plot (in the observations).

The figure was chosen to be comparable with Søvde et al. (2008), but we have included a figure (b) focusing on Europe. The improvement to Søvde et al. (2008) is mainly in north-east of Europe, towards Siberia.

Figure 6 is almost unreadable. Too small, and labels are not resolved and very difficult to read even if enlarged.

The figure was produced for A4 format, and we have checked its readability on paper. In addition, it is possible to zoom in on the figure (although figure quality may perhaps depend on the pdf reader).
However, we have split the figure in two to make each profile larger and more readable.

Figure 9 – a few random sondes don't reveal much about the model. Do you have chance of a sequence where you could see some feature? I don't see how this figure supports statements in the text about folds.

Tropopause folds are shown in figure 9a and 9d, and to some extent in 9c. However, we have taken the suggestion from reviewer #2 and changed this figure with a CO comparison. In addition, we have produced a supplementary document where the models are compared to all single sondes used in the monthly means. The discussion on folds have been kept, but changed to take this modification into account.

Figure 10 – also small, hard to read and pretty much impossible to interpret.

We have made it more readable, checked on print-out and on screen.

Some of the grammar problems
1563 Line 2 CTM2 resulted from (rather than 'was derived from')
Done

Line 7 better transport and diagnostics have been added – the core CTM does not 'develop'
Done

1564 L 25 – no need for last word 'applied'.
Done

1578 L6 delete the comma after available.
Done

1568 L 21 no information . . .is used . . .
Done

1569 L 25 'and therefore treats' – grammar issue. 'splitting it' – not clear what 'it' refers to.
'and therefore treats' changed to 'treating', and 'splitting it' changed to 'splitting this precipitiation'

1575 L 24 The C3 simulation produces
Done

1578 L 4 – the models produce reasonable polar O3 loss – making a statement, rather than the weaker 'models are capable of' (the model actually has no choice).
Thanks for pointing this out, we have changed the text.

1578 L16 differences are seen
Done

1578 L 22 The CTM2 produces
Done

1578 L 26 delete ‘also’
Done

1579 L 4 CTM3 reproduces
Done

1594 L 3 somewhat larger? greater? Improvement
Thanks for pointing this out. Changed to “closer to 50%”.

1595 L23 mainly increases computing time.
Done

1596 L 13 profiles place (rather than places)
Done

1596 L 23 giving a negligible lower tropospheric O3 burden (as opposed to negligibly)
We have rewritten this sentence, as it could be misleading. The effect discussed was to give a slightly lower burden of tropospheric O3, not a negligible lower-tropospheric burden. The new sentence is “… giving a negligibly lower burden of tropospheric O3”.

1597 L 17 slightly less of e.g. HNO3 – what is HNO3 meant to be an example of?
Changed to “removes slightly less of soluble species from the troposphere”.