Replies to Referee #2: Michelle Santee

It is not clear what the point of such a detailed comparison with observations really is. Obviously the CMAM30 fields lack the fidelity needed to be useful for specific polar processing studies (for which a CTM is probably a better choice at this point). Is the ultimate goal to be able to employ the nudged CCM in investigations of polar processing and ozone loss during particular winters? Or is it to enhance CMAM's polar processing capabilities for studies of future climate? Towards the end of the manuscript, it is mentioned that a newer version of CMAM has been developed that includes a more realistic treatment of chlorine activation. Will the results found here inform further model refinements? Are there plans to generate another 30-year data set with the improved CMAM? If so, it might have been more illuminating to perform these kinds of comparisons with a revised data set.
CMAM30 is a publically available dataset and should therefore be validated. The idea is to allow users to know where and for what time periods the model data provide a reasonable depiction of the real atmosphere (in terms of tracers). It is agreed that a comparison of a CTM with a detailed representation of PSC microphysics would allow a much better assessment of our understanding of polar processes and our ability to represent them in a model. We note that the representation of stratospheric processes in the model used for the CMAM30 simulation is the exact same as has been used for free-running, century-long simulations that project the future evolution of the ozone layer for the Chemistry-Climate Model Validation (CCMVal) project and the WMO/UNEP Ozone Assessments. By constraining the dynamics to observations through nudging, we are able to directly compare the model chemical processes to satellite observations to gain a better understanding of how the present formulation of CMAM is able to capture the processes that give rise of ozone destruction. The improved version of CMAM mentioned in the article has, in fact, been significantly motivated by the comparisons presented here. Given the significant effort required to produce the CMAM30 dataset, it has not been possible to redo the CMAM30 simulation and the comparisons presented here. There are plans for a 'version 2.0' of the CMAM30 dataset in the near future, with the present manuscript representing our assessment of the first version. A statement in the second last paragraph about further model development has been added.

Specific Substantive Comments:
P11183, L25-28; p11184, L1-2: It is stated that: "PSCs form when the temperature dips below 196 K for Type 1 PSCs, which are composed of nitric acid, sulphuric acid and water, and below 188 K for Type 2 PSCs, which are solid water." This appears to be a general statement, and not a description of CMAM's treatment of PSCs. It is too simplistic, however, as the PSC formation thresholds are not constant values, but depend on pressure and HNO_3 and H_2O abundances. In addition, PSCs are important for ozone depletion not only "during spring", but throughout the winter. Finally, "sedimentation of the heavier ice particles" leads to the question "heavier than what?" It is probably better to just say "large particles".
This is meant to be a general statement. "in the stratosphere" has been added to convey this. The statement has been changed to state that the temperature thresholds are not the only factors (although no specifics are given). "Heavier" is changed to "large".
P11186, L11: Carleer et al. [2008] is a rejected paper and should not be cited. I understand that it is "published" in ACPD, but since it never completed the peer-review process, it is not a valid reference. Papers submitted to but not accepted in JGR (for example) are not citable, and nor should rejected ACPD papers be.
Removed. Hegglin et al, 2013 (the SPARC Data Initiative paper on water vapour) has been referred to, although it is not a validation paper (noted in the text that this is a climatological approach to validation).
P11186, L16: A reference is needed for sPV, since not everyone is familiar with this concept, and

moreover there are various ways of scaling PV. The Manney et al. [2007] paper cited for the DMPs would probably suffice.
Cited.
P11187, L1-4: There are two Froidevaux et al. [2008] validation papers – one for HCl, the other for ozone. Both need to be cited here (as they were in the originally submitted manuscript), but in the current draft only the HCl paper appears in the references. In addition, because it contains updated information on the v3 data used in this study, it would be appropriate to cite the MLS Quality Document: Livesey, N.J., et al. [2013], Earth Observing System (EOS) Aura Microwave Limb Sounder (MLS) Version 3.3 and 3.4 Level 2 data quality and description document, JPL D-33509 (available from the MLS web site).
Both Froidevaux et al papers are now in the document. Livesey et al., is referred to.
P11189, L14: The statement is made that OSIRIS temperatures have not been validated near 85 km. But this is not true – according to P11185, L17-19: "temperatures between 80 and 86 km are deemed unreliable, due to uncertainties in O2 absorption in the A-band. The temperature data also exhibit a cold bias of 10–15 K near 85 km (Sheese et al., 2011)". Thus the data have been validated, in that their quality has been evaluated and documented.
Changed to "...are known to be problematic."
P11190, L6-11: It is stated that "Agreement with ozone seems to be better", but this is somewhat difficult to judge quantitatively since the temperature comparisons are described in terms of absolute differences (K), whereas the ozone comparisons are described in terms of percent differences. I'm not convinced that the ozone comparisons are that much better, at least in the UTLS. The authors point out that small ozone mixing ratios in some regions lead to large relative differences, which brings up the question of why it was felt that relative rather than absolute differences were better to show for ozone.
It is standard to use relative differences for trace species. Here, this done in order to be able to compare with other validation studies. Temperatures were compared with absolute differences for the same reason. Given previous studies, this seems a reasonable approach. The fact that the agreement in the UTLS is not as good is discussed and possible reasons are presented. The subjective statement about comparisons with ozone being better has been removed.
P11190, L18-20: "Note that compared to ACE-FTS, CMAM30 shows a low bias in the upper troposphere/ lower stratosphere in the tropics (see Fig. 4), in contrast to the OSIRIS data." I don't believe that the contrast with OSIRIS is that strong – Figs. 2b, c, and d all show a low bias in CMAM30 relative to OSIRIS in the tropical UTLS.
This statement has been changed to point out that the April and October biases with respect to ACE-FTS are much larger and extend over a larger region, particularly for October. Note that the low bias relative to OSIRIS in October is very small and occurs in one latitude bin. The bias in April compared to OSIRIS is not as large and not as deep as that relative to OSIRIS. Nothing can be said about July.
P11190, L22-25: "At 0.1 hPa, the diurnal cycle of ozone becomes important in the comparisons, and while ACE-FTS ozone profiles extend to this altitude, they have not been validated above 70 km". Given the importance of the diurnal cycle and the fact that the reliability of the ACE data has not been demonstrated at this altitude, what can we learn from the model/measurements comparisons there?
ACE data have been validated up to 70km (~0.1hPa for the most part) and so they is compared. If the question is about exploring above that altitude, then this question is beyond the scope of the paper.
P11191, L1-8: A small positive bias is also found in the tropical lower stratosphere in October. It's a small feature, but it stands out against the otherwise nearly pervasive low bias.
This will be explored in an upcoming paper, and not discussed here. The upcoming paper is already mentioned elsewhere in the paper.
P11191, L9-10: I don't think it's quite fair to characterize methane as having an overall high bias

of 10–20% when so much of the vertical range (especially at high latitudes) in most months in Fig. 6 shows biases of 60–100%.
Changed to "over much of the stratosphere" and "a background bias", which was what was meant by overall in this case.
P11191, L13-17: These sentences require some clarification. I assume that the statement "The methane profiles exhibit a kink between 10 and 1 hPa, which descends over a season" refers to the ACE-FTS data and not the model, but this needs to be made explicitly clear. Is this kink in the observed methane profiles a real atmospheric feature? Some information on the validation of the ACE CH_4 measurements, in particular for this kink in the profile, should be included.
"…in both model and observations…" has been added to clarify. "over a season" has been changed to "over a few months". A reference to de Maziere et al. 2008 has been added to indicate that the kink is a real feature in different satellite datasets.
P11192, L2-4: "It is clear that the CMAM30 data set does not simulate enough ozone destruction during this period". Couldn't it also be that the diabatic descent in the model is too strong, bringing down too much ozone-rich air from above? In L20-23 the possibility that horizontal mixing out of the lower stratospheric vortex is underestimated is mentioned, but the strength of the modeled descent does not seem to be explored. The methane comparisons could be used to explicitly discuss the reliability of the vertical transport in the model. Finally, it is asserted in L25-26 that the model methane indicates that "the polar vortex isolation is not the answer to this puzzle". Please expand on why this is the case. What specific aspects of the methane distribution preclude this explanation?
The diabatic descent was inferred in the statement on transport, so "and too much diabatic descent within the vortex" has been added to the text. The fact that there is no build up in methane and that the methane "kink" descends at the same rate in the model and in the observations suggest that the descent in this region is reasonable.
P11193, L4: This sentence is a little confusing – it's not clear exactly what is being compared. I believe that the authors intend to compare the growth in CMAM ozone biases between July and September to the growth in CMAM methane biases over that interval, but I'm not entirely sure – as written the sentence seems to imply that the growth in ozone biases is being compared to the methane biases themselves (and not the growth therein). Please clarify.
The point here is that the differences in the growth between ozone and methane means that the ozone bias by the end of the season is likely due to not enough chemical loss, rather than too much transport. "..growth in.." methane biases was added.
P11194, L19-20: It is stated that 2006 was chosen as a "typical year", but in fact 2006 was characterized by a record-setting ozone hole. It no doubt represents an excellent year to compare against model results, but I'm not sure that a record ozone hole year can be called "typical" even in the SH.
"Typical" is meant to refer to the fact that the discrepancies between the observations and the model results are similar from year to year. The text is changed to reflect that.
P11194, L22-23: Are the ozone discrepancies really greatest in the 2004/2005 NH winter? The MAD is largest that year, but the percentage differences at the lowest level in January 2007 and 2008 appear to be just as large as those in January 2005, if not larger.
For the winter seasons, January 2005 has large and variable difference between the model and ACE-FTS and these differences have more vertical depth than the other winters. The largest discrepancies are actually in summer 2008.
P11194, L23-24: There is still not enough information given about how the error bars on the ACE and MLS data points are calculated. It is stated that: "Instrument errors are calculated according to the reported errors for each profile and averaged using the square root of the sum of squares of the errors". But it is not clear whether the "reported" errors include accuracy or just precision. It is also still not clear why some MLS points appear to have errors much larger than those of surrounding points. Since it is the relative changes and their comparison to those in the model that

are of most interest here, I believe that it is not so important to include accuracy in the error bars – it is the precision term that really shows how well the instruments can track day-to-day variations in the atmosphere. The approach taken by the authors fails to account for the fact that precision (which reflects radiance noise) can be greatly improved by averaging. ACE-FTS measures at most 15 profiles inside the vortex on any given day; on some days no more than 3 or 4 profiles may be obtained in the vortex. In contrast, MLS measures many profiles inside the vortex on almost all days. The "precision uncertainty" on the MLS vortex averages is probably reduced over the single-profile precision by at least a factor of 10 in most cases. The error bars on the plots in Figure 13 should be recalculated for both instruments to reflect only the precision, divided by root N (where N is the number of points contributing to the vortex averages in each case).

Unfortunately, there was an error in the original manuscript. Plotted errors were cumulative in order to provide an envelop for possible values. Now, both cumulative and mean squared errors are shown for Aura-MLS. Only cumulative errors are shown for ACE-FTS, although the reported errors for ACE-FTS are small so the cumulative errors are similarly small. In addition, more discussion on the error analysis, and the motivation for it has been added in the text. The fact that ACE-FTS is only sampling a small region of the vortex each day is clearly pointed out, and demonstrated in the figure, and the discussion is mindful of the limitations of the nature of the orbit. Figure caption 13 has been updated with the new error bars.

Figure 13: In addition to the error bars, there is another substantive issue with this figure: the y-axes for the two hemispheres in the middle and bottom panels are inconsistent. In the middle panels, the left-hand y-axis for both hemispheres represents water vapour. Only the SH left-hand y-axis is labeled, so the reader expects the tick marks on the NH left-hand y-axis to be the same as those of the SH left-hand y-axis, but they are not. Therefore it is not possible to derive quantitative information about NH water vapour from this plot. The rightmost y-axis shows HNO3 amounts for both hemispheres. But it would be better if, rather than appending another vertical axis at the side of the plot, the right-hand y-axes of the panels for both the NH and SH showed the HNO3 range (only the NH axis would have to be labeled, just as only the SH H2O axis is labeled). Similar issues are present with the left-hand y-axes of the bottom panels – that is, the O3 / HCl/ ClONO2 range on the left-hand y-axis for the NH in the right column does not align with that of the SH in the left column. Again, the ClO range should be presented on the right-hand y-axes of both hemispheres.

More axes have been added in colours corresponding to the curves. The HCl and ClONO2 axis has been added in the bottom panels, although the ClO axis is left only on the right panel since it seemed impractical to add that to the left panel.

P11195, L1-4: I do not believe that it is adequate to cite "technical documents available from the Aura-MLS website and the ACE-FTS website". I would think that a general reference such as Bernath et al. [2005] or Boone et al. [2005] or [2013] would suffice for ACE-FTS, and the Livesey et al. [2013] quality document would be good for MLS. In addition, I think it would be better to label this quantity something other than "area of the polar vortex", which most readers will interpret to mean the true area encompassed by the vortex as defined through meteorological analyses, not the "area of the vortex sampled by the satellite instruments on any given day".

None of the published papers include the horizontal resolution or footprint. This information is included in Livesey et al. (2013) so the reference is included.

Discussion of Figure 13: More generally, do the top panels of Figure 13 imply that ACE-FTS samples virtually nothing of the polar vortex in either hemisphere? The blue triangles deviate very little from zero in either panel. This may make sense, as ACEFTS obtains only a few profiles inside the vortex on any given day (as illustrated in Figures 11, 12, and 14). It does, however, call into question the value of the ACE comparisons in these figures. Given the difference between the area of the vortex sampled by MLS and that sampled by ACE-FTS, do the ACE comparisons really add very much information? Finally, what level of the atmosphere does

Figure 13 pertain to? Does it also show results for 500 K, as do Figures 11, 12, and 14? My apologies if I missed this information in either the main text or the figure caption.
ACE-FTS does sample the polar vortex, but the actual area is small. Note that the values are not zero. ACE-FTS comparisons are useful for ClONO2, and for independent data. The nature of the orbit must be accounted for when interpreting the data, and this is explained in the text. "500K isentropic surface" has been added in the text and in the figure caption.
P11195, L12: CMAM is only higher than ACE-FTS water vapour in midwinter – it agrees well with both ACE and MLS at the end of winter.
This is a problem in itself since the rest of the stratosphere shows an overall low bias of 10-20%. This means that although the water vapour agrees, it doesn't agree for the right reasons and so shouldn't be trusted. This point is made in the text by adding the statement "Note that although the H2O amounts appear to agree, CMAM30 should have a ubiquitous low bias to be consistent with earlier results. Although the H2O agrees in this region, it is because there is not enough water vapour at the beginning of the season, and not enough is sequestered as PSC aerosol."
P11196, L6-12: Interpretation of comparisons with ACE-FTS data in September can be confounded by the ACE sampling, which sweeps rapidly through the collar region during this period. As shown by Santee et al. [JGR 113, D12307, 2008 (not the ClO validation paper)], it is difficult to separate increases in ACE-FTS ClONO2 and HCl arising from abrupt changes in the air masses being measured from those arising from chlorine deactivation. Thus caution is required in drawing conclusions about model performance based on these data, and this complication should be mentioned.
This change in orbit is already discussed in the paper. The reference has been added and the statement "The fact that ACE-FTS samples only certain latitudes on certain dates must be taken into account when comparing the data" has been added.
P11196, L27-28 to P11197, L4: Although the model does not account for NAT PSC formation, it does include a treatment of STS. Therefore should it not show some signs of PSC formation in this cold winter, during which extensive PSC formation has been documented? I'm not sure that I agree that "The gas phase HNO3 for the NH polar vortex also agrees well over the whole season" (agrees well with what?). It is too high compared to MLS until mid-March, after which it is considerably low relative to both MLS and ACE.
The text is changed to state that CMAM30 HNO3 agrees reasonably well with AuraMLS over the winter season, but is too low in spring compared to both AuraMLS and ACE-FTS. It is difficult to make an assessment with ACE-FTS since there are few measurements over the winter (in January, the orbit is such that only the pole is measured, skewing the measurements. Certainly the HNO3 agrees much better in the NH winter polar vortex compared to the SH vortex. The statement that "...otherwise there are no signs of PSC formation in the model.." has been qualified by "(i.e. solid phase H2O is absent)" to clarify.
P11197, L8-9: The authors state that the "The seasonal variation in HCl from Aura-MLS and ACE-FTS agrees well over the season", but they might note that the degree of agreement between MLS and ACE HCl measurements over this particular winter was discussed in detail by Santee et al. [JGR 113, D12307, 2008].
The reference has been added (and text to support the reference).
P11197, L14: As noted elsewhere and illustrated in several of the figures in this paper, a "vortex average" for ACE is not the same as a vortex average for MLS or the model. The ACE averages do not encompass the same air masses, and this likely accounts for the much greater degree of variability in the ACE results. To reduce the day-today variability in this plot, the authors might consider imposing a minimum number of profiles required to define a "vortex average" and then discarding averages failing to meet this criterion.
Given the limited area that ACE-FTS samples and the zonal nature of the vortex, taking out some of the points is unlikely to change the results. The real issue is not the day-to-day variability but the fact that the instrument is only measuring at certain latitudes on certain days, and the latitudes

change over the season. ACE-FTS results were not removed.

P11199, L13-18: These sentences need to be worded more carefully. Sedimentation of PSC processes is not included in the model. Thus it is not precisely correct to state that the treatment of PSCs “does not allow for enough denitrification in the lower stratosphere” – it doesn’t allow for *any* denitrification (which means the permanent removal of HNO3 from the stratosphere through the settling of PSC particles). The statement “In addition, the dehydration of the lower stratospheric vortex in the model does not seem to be enough during the winter months” is similarly incorrect. It isn’t a question of degree – the model simply cannot simulate dehydration, period. In addition, it was stated in Section 2 that the formation of STS in the model is calculated following Carslaw et al. [1995] and that of water ice PSCs is based on supersaturation with respect to the frost point. Therefore, presumably PSC formation is not “turned on” in the model exactly at the thresholds of 196 and 188 K, as implied here.

The model can temporarily dehydrate/denitrify the air in the vortex, but not permanently. The text is changed to clarify this, specifying that it is the denitrification in the spring that is important. The wording is changed to reflect that the temperature thresholds are approximate, rather than exact.

P11200, L1-4: “In CMAM, as the winter progresses this pathway to chlorine activation shuts down once the ClONO2 is depleted”. Does the ClONO2 in CMAM become completely depleted? It is quite low at the beginning of winter, but Figure 13 does not seem to show complete depletion of modeled ClONO2. Of course, further conversion of reservoir chlorine into active forms can occur later in the winter when photolytic processes reform ClONO2.

We find that ClONO2 does completely deplete at latitudes where there is complete darkness. As the partitioning of reactive chlorine to ClONO2 is not advected, the depletion lasts until the sun rises at that latitude, at which point photochemical production of additional ClONO2 allows for further processing. For example, the small concentrations of ClONO2 in July shown on Figure 13, being the vortex average, is a combination of practically zero values in the dark portions of the vortex and higher values at lower latitudes. “…where there is complete darkness…” has been added to clarify

P11200, L20-21: “Even during a cold year without an SSW, water vapour shows very little change from the consistent low bias”. But this is not surprising. Although 2004/2005 was a moderately cold winter, temperatures were not exceptionally low and only a single event of depleted water vapour at the end of January was linked to formation of water ice PSCs [Jimenez et al., GRL, 2006], thus essentially no dehydration occurred that winter.

The reference has been noted in the text. Unfortunately, the exceptionally cold year of 2011 is not included in the dataset. Jimenez et al. (2006) do note that the dehydration that does occur and it does affect the stratosphere several months later.

Typos and Other Minor Wording / Grammar Comments:

P11184, L12: denitrification/dehydration : : : *do* not occur

The point here is that they **do** occur in the real atmosphere. The next paragraph then contrasts this with what goes on the model. A qualifier “in the stratosphere,” has been added to clarify what is meant.

P11185, L20: “multiple scatter” –> “multiple scattering”

Changed as noted.

P11186, L16, and p11187, L4: I don’t think that “DMPs taken from the GEOS5 model” or “DMPs taken from MERRA” is the best wording. The information is not simply copied from the analyses. Rather than “taken from”, I think that “calculated from” or “based on” (or, of course, “derived from”) would be better. In addition, GEOS5 is not purely a model. It would therefore be better to say “the GEOS5 data assimilation system”.

Changed to “calculated from”. “GEOS5 data assimilation system” has been added.

P11186, L27: delete comma after “2005”

Changed as noted.

P11187, L1: “found *in* Froidevaux”
Changed as noted.
P11188, L6: “SD” might be a standard acronym, but it should still be defined the first time it is used.
“SD” has been removed and replaced by “standard deviation”.
P11189, L13: It would be helpful to add “(see Fig. 3)” at the end of the sentence about ACE temperature comparisons.
Added.
P11192, L12: This wording is ambiguous for the Arctic. I assume that the years given refer to January dates, but it would be clearer to write them as “2004/2005”, etc.
Changed as noted.
P11192, L16: It would be appropriate to include a general reference for the statement about the strong and chemically isolated SH polar vortex, at least the WMO Report if nothing else.
Added.
P11192, L20: Dehydration (in the sense of permanent removal of water vapour from the stratosphere) requires not only PSC formation but also particle sedimentation.
Temporary dehydration can occur while the model sequesters water vapour as solid. The point here is that the model may not have enough of this temporary dehydration. This is clarified in the text.
P11192, L27: “comparison : : : *shows*”
Changed as noted.
P11193, L5-6: It seems slightly jarring to introduce Figures 11 and 12 after they have already been discussed in the previous paragraph (L2).
Added “(see Fig. 11 and Fig. 12 below)” to indicate that they will be discussed following.
P11193, L17-28: It would be better to differentiate between the light and dark greys in this sentence.
Added to the text in the brackets.
P11193, L25: “cycles : : : become” –> “cycle : : : becomes”
Changed as noted.
P11194, L9: What exactly does “PSC concentration” mean? Are these plots of aerosol extinction coefficient, or just frequency of PSC occurrence?
“PSC concentration” is replaced by “solid H2O” to be more precise.
P11195, L18: “HNO3 gas phase” is slightly awkward and inconsistent with usage elsewhere in the manuscript, where typically “gas phase HNO3” is used.
Change as noted.
P11196, L1: This wording about ClO seems to imply that it destroys ozone only in darkness, which is not true. Sunlight is required for significant ozone destruction in the polar lower stratosphere, and such chemical loss starts at the sunlit edge of the vortex in June and sweeps poleward with the terminator as winter progresses. I do not see why such a statement is even necessary at this point in the manuscript. It is sufficient to state that active chlorine is necessary for chemical ozone loss and CMAM produces too little of it.
Removed “and can destroy ozone during the dark winter months”.
P11196, L4: delete the comma after “bite”
Changed as noted.
P11196, L15-16: A general reference for SSWs in the NH would be appropriate.
Reference to Andrews et al. 1987 is added.
P111978, L7: “the ozone maximum (_1 hPa of _8 ppmv)”. This wording is confusing. It would be clearer to say “the ozone maximum (_8 ppmv at _1 hPa).
Changed as noted.
P11198, L23-26: It is unnecessarily redundant to say “: : : may be explained by a Brewer–Dobson circulation in the model that is too fast, : : : as air parcels circulate via the Brewer–Dobson

circulation". I suggest deleting the phrase "via the Brewer–Dobson circulation" at the end of this sentence.
Changed as noted.
P11210, Fig. 3 caption: Is this a typo, or does panel (b) really represent a particular day in April?
"9" has been deleted.
P11218, Fig. 11 caption: The white overlaid (temperature) contours need to be defined in the figure caption (not just in the text).
This hasn't been added in the figure since the figures are already quite busy.
P11220, Fig. 13 caption: In the penultimate line, delete the comma after "HNO3" and add one after "panels". In addition, it seems awkward to write "sPV <> _1.2PVU", and it may not be clear to all readers that these values refer to different hemispheres. Wouldn't it be easier to use absolute values? Finally, it is very hard to read the x-axis labels in these panels. I suggest using a smaller font so that the date labels do not run together.
Comment on sPV has been changed according to referee #1. A second legend for the SH has been added and the axis labels have been scaled to (hopefully) be more legible.