Reply to the review from Referee 1

We are thankful to this referee for the review and the associated suggestions, listed in italics below. We provide our detailed responses (regular font) and plans; our revised manuscript will be available in a fairly short time.

It would seem from the referee comments that there are no demonstrable big issues with the science (or math), besides some requested clarifications, and we are pleased that Referee 1 found our manuscript to contain "some interesting results" [and Referee 2 found "a lot of valuable and detailed information"]. We hope, furthermore, that the plans we describe herein for clearer messages and revisions will be of a satisfactory enough nature, or we will need to ask for more specific comments.

(1) This paper evaluates two versions of the WACCM model using satellite observations, mainly from Aura MLS, but also using multi-instrument compilations. The paper contains some interesting results, but it is also very long (70 pages in the submitted format and 32 figures, plus supplement) and focused on one specific model.

Reply: We are planning to cut down on the length of this manuscript, mainly by relegating some of the less critical Figures to the Supplement. Although this does not necessarily translate into a very large cut in terms of text length, we consider this work to be a fairly comprehensive analysis, which therefore leads to a longer paper; there have definitely been some longer (atmospheric) papers in the literature, and specifically in ACP. Turning this into two separate papers mainly for the sake of overall length seems too artificial, and this would be quite an elaborate proposition, with the need for some duplication regarding both the data sets and the models; as an aside, this would actually lead to more reviewing work for the community. We hope to have shown that detailed analyses are necessary to enable identification of both good agreement (a result in itself) or significant differences between model runs and the data sets, but also for some of the more subtle differences, and furthermore, that an understanding and discussion of error bars and potential data issues is important. We will also strive to reduce the amount of text in the revised manuscript, especially where some less critical aspects can be discussed more succinctly, or taken out altogether. In particular, we plan to shorten Section 5.1.1 (pages 11-13) to a text length that roughly matches (rather than exceeds) the text length of Section 5.1.2 (on variability issues); the cuts to Sect. 5.1.1 will be of order 30% (or more).

In terms of reducing the amount of Figures and related changes, our specific plans are to remove Figs. 13, 14, 15, and 17 from the main text (and relegate these to the Supplement, with a slightly shortened discussion), since these mainly reinforce the expectation (already noted for O_3 and H_2O) of better model/data fits from SD-WACCM, as one might expect from a model with better dynamical constraints than the FR-WACCM version. Such an expectation does not hold for the variability diagnostics, so these are really best left in the main text, although we will plan to displace Figure 22 (on the N_2O and HNO_3 variability comparisons), and move it to the Supplement. Moreover, we feel that Figure 31 on lat/p contours of short-term trend for various species can be moved to the Supplement, as it is less critical, and given past (and ongoing) work on this topic. While Figure 32 is interesting to us, it is more of a side note on lower stratospheric tropical cohesiveness for various species exhibiting similar dynamical variability, so we decided that the text and Figure in this case can be eliminated altogether without much of an impact on this paper.

In summary, the total number of Figures in the main text will be trimmed down by almost a quarter, with a more manageable total of 25 Figures; writing up a multi-year effort of (part-time) work on detailed model/data comparisons is bound to lead to a longer manuscript than several shorter analyses; to our knowledge, fits, correlations, variability, and trend comparisons are rarely investigated to this extent in model/data comparisons, even for a single model (or two flavors of one model). This, with some reductions (and clarifications) in the text (including the Abstract and Conclusions section), will at least show our good faith effort towards the referee comments. Recommending a goal of exactly 20 Figs. (as done by Referee 2) is rather arbitrary, but our point here is that we have considered these requests with some care, and that we are being responsive.

(2) The paper contains an evaluation of the model (SD and FR) for 5 species compared to satellite observations. The paper points out general agreement and areas of disagreement, but the reasons for any disagreement are not really looked into (except for HNO_3 and the lack of ion chemistry).

Reply: There are several aspects to these model/data comparisons. Looking carefully into reasons for disagreement can be the subject of separate papers altogether, possibly involving new model runs (which would take quite a bit of time), and this would also increase the length of this (already long) manuscript. We will point to where some likely causes can be mentioned, although in all fairness, we believe that this has already been done in several places (see more in the numbered list below), and beyond the HNO_3 issue mentioned by this referee. However, we are also adding more information and discussion in various places (see further below).

In particular, we provide further explanatory material in some of the following areas, ignoring from this list the HNO_3 issues (and lack of full model chemical pathways) already sufficiently described in the manuscript. Without the exact revised text for now, please see the following list of specifics, although some of the items in this list are there to provide some rebuttal to implications that we provide few explanations besides showing the comparisons themselves (or the advantages of one model version versus the other). We strongly believe that these comparisons (in themselves) are worth displaying in a publication, even if this only applies to the WACCM model, which is considered state-of-the art. Moreover, and almost as importantly, we have shown that areas of disagreement very often fall within the estimated error bars, so there are not that many really significant discrepancies; we hope to take some credit, in fact, for trying to be careful about including realistic error bars in many of these comparisons. However, an investigation into other models for similar areas of agreement or disagreement is beyond the scope of this work, which we consider a first step that can help other modeling groups focus on certain regions of potential disagreement. Later on, another paper could hopefully identify where, and maybe why, certain models do better than others in certain places or time periods; in fact, some of this may already be "in the works" or near completion (based on a list of planned studies for CCMI at www.met.reading.ac.uk/%7Eqr903932/CCMIwebsite/Wordpress_PDFs/CCMI1_PlannedAnalysis_20170715.pdf).

1. *Page 6, 1st paragraph*: We now make the point regarding the Fig. 2 (and Fig. 1) model/data lower stratospheric O_3 differences near 50°N-60°N (even if it may be obvious) that transport-related model issues (not chemistry issues) are the most likely reason for the models to significantly overestimate mean ozone and its seasonal cycle at mid- to high latitudes. In addition, we are adding related information in the text for H_2O comparisons, given that we also see a significant (factor of two) WACCM model overestimate of the MLS H_2O fields (mean value and seasonal amplitude) in the same region (detailed plots not shown); this discrepancy goes beyond a (previously documented) 30-40% dry bias of MLS H_2O versus sonde data a few km below the tropopause. However, digging into model details (or even the meteorological fields),

in addition to possible other data sources (or data issues) for O_3 (and H_2O) comparisons would need to be the subject of a new investigation, interesting as it might be.

2. *Page 8, lines 6-14, and Fig. 7*: Regarding the seasonal changes over Antarctica, our analyses include species other than HCl and provide more of a climatological description regarding this discrepancy in HCl behavior than what was shown in the paper by Groos et al, (ACP, 2018). The latter work attempted to ascribe such a discrepancy to various factors, without a fully satisfactory answer, and we do not currently have further thoughts on this topic, as more detailed investigations (not speculation) would be required to make further progress. On the same topic, we do provide a likely explanation for the better matches from SD-WACCM (vs FR-WACCM), namely the connection to more realistic temperatures.
3. *Page 11, lines 36-38 and Section 5.1.1*: The better SD-WACCM results (here and in this section more generally) regarding model/data fit diagnostics and model/data correlation coefficients are related to the better dynamical description for the "specified-dynamics" version of WACCM, as we point out in this section in more than one place (see also the 2nd part of the top paragraph on page 12, regarding H_2O comparisons). This is the main result from the discussions on pages 11-13, and a result that is worth including in this paper (in our opinion), even if there are other (probably more illuminating) results.
4. *Section 5.1.2 on model/data variability comparisons*: The interannual variability in monthly means represents a useful diagnostic of model/data comparisons, and it also relates to trends and detectability of trends, as we point out in this section. The main variability disagreement between models and data involves water vapor, a species that is also more difficult to model, given its different phases and its more complex pathways for entry into the stratosphere, the influence of ENSO and cold point temperatures, as well as the QBO and circulation changes, along with changes in methane and (mainly in the mesosphere) the solar cycle impact. Some of these processes (or their variability at least) are possibly not sufficiently well represented in either SD-WACCM or FR-WACCM, but there are better fits to the data from SD-WACCM. Also, on the variability issue, we do make the point that the H_2O interannual variability is underestimated by the models, and since the uncertainty in trend detection

depends on the variability, a larger than modeled atmospheric variability implies that it will take longer than expected to detect long-term trends in water vapor.

5. *Section 5.2 (trend comparisons)*: Here are the main points for each species:
   - **O_3:** MLS data alone have not been used yet to document trends (for the MLS years of operation that overlap the model runs), so this is a novel result, even if the time period is short enough that the (2σ) trend error bars are often fairly large. If one averages the results over the upper stratosphere, there are robust indications of an increase, based on the MLS data alone, and this avoids some issues associated with merged data sets (e.g., changes in spatio-temporal coverage between different instruments). It is also interesting to see indications of increases in the tropical lower stratosphere (albeit with less robustness than in the upper stratosphere), in apparent agreement with the SD-WACCM results. Most notably, the lat,/height patterns of trends, ignoring the absolute error bars, are remarkably similar for MLS data and SD-WACCM (see Fig. 25); we feel that this is a very informative plot. Furthermore, version 2.20 of GOZCARDS O_3 is evaluated for trends in this work, and we highlight some differences versus the original GOZCARDS data set.
   - **H_2O:** The main points for this species are now made more relevant, we hope, in the context of what one might expect from longer-term trends versus what happened during the shorter-term (2005-2014) versus MLS trends, which are significantly larger than what one would expect from the water vapor changes caused by increases in methane alone. We also note that FR-WACCM trend results are significantly smaller than SD-WACCM (and observations), but this does not imply a longer-term systematic underestimate from FR-WACCM, based also on our looks at longer-term time series (although these are not displayed specifically in this manuscript). The Abstract has now been changed to reflect these points as well.
   - **HCl:** We are not planning much change regarding these results, and we think that the main points are clear enough: there is some underestimation of MLS HCl trends from the models, and some LS tropical positive trends in these observations which deserve further investigation. However, we will add a pointer to recent work (if it gets in press soon) that shows the impact on HCl trends of a better treatment of VSLS and their trends, as this seems to be a way to close at least part of the gap

(model versus data trends). The other issue could be related to an MLS overestimate of the HCl LS trends (as this is what happens in the upper stratosphere, a known issue for MLS HCl there).

- **N_2O and HNO_3:** There is good agreement for these two species overall, in terms of the model versus data trends. Some of these trend variations versus height (particularly for N_2O) must be related to stratospheric age of air and circulation, but we also clearly see (in time series not shown here) that the QBO, in particular, has a large impact on the variability, as one moves away from the tropopause region; this is a well-known feature. This large percent variability (as one reaches the mid-stratosphere) swamps the underlying long-term trends in N_2O. The WACCM time series capture the observed (MLS) variability remarkably well, and the trends for 2005-2014 reflect this sort of agreement (Fig. 30). There are some slightly larger differences in terms of the somewhat poorer phasing of variability (and fits to the data) for FR-WACCM, but the main features versus latitude and height are well reproduced. This also holds for HNO_3. We will thus add a few words very similar to these in this part of Section 5.2, in terms of our understanding (and at least partial explanation) of these trends and their variations.

Our draft revised Abstract (see below) also hopefully clarifies the main points in a somewhat better way (without making it much longer), as a response to the Referee comments. The revised text will add related information for clarifications and context; it will also be trimmed elsewhere to try to address the issue of paper length inasmuch as possible (without losing too much content).

(3) The comparison of the performance of the SD and FR models is a main focus of the paper. There are differences but overall conclusions on the accuracy of SD models, for example, seem to be missing.

Reply: We do not fully understand this comment, but we will attempt a reply that covers the options. It is <u>really</u> beyond the scope of this work to try to dive into why SD models differ from one another, if that is the reviewer's point, although we think that this would be an interesting study for the future. We have examined only the SD-WACCM/MERRA model in detail in this study using multiple diagnostics. There will be future papers that compare processes and biases between the participating CCMI SD models (as mentioned earlier). The point of our work is to perform a detailed model/observational analysis of two configurations (FR and SD) using the same modeling system (i.e., CESM framework). Here, the tracer advection routine (Lin, flux form finite volume) and WACCM chemistry module (for gas-phase, heterogeneous, and photolysis reactions) are identical between the two configurations. The differences between the two configurations are mainly due to how the circulation is derived. The FR configuration allows the ozone to be interactive with the heating rates and therefore circulation. The SD configuration uses a specified meteorology that drives the circulation. Therefore, when we compare the FR and SD model versions observation-based diagnostic, the "goodness" of the results between FR and SD removes uncertainty of both the advection and chemistry assumptions (since they are the same). However, there is still uncertainty in the derivation of the circulation in FR and the nudging approach used with the observed meteorology. The approach described in this paper is essentially a first step in understanding how well models represent biases and variability in comparison to observations. The next step could be, of course, to examine diagnostics across multiple model systems, but not here (see also our response to item (1)). We plan to change part of the Introduction (and maybe the model section also) in the revised version to better motivate the purpose of our analyses of FR-WACCM versus SD-WACCM, as mentioned above.

If, on the other hand, the reviewer is asking about the accuracy of the specific SD-WACCM model run used here, most of the comparisons here show that there are few large areas of disagreement, beyond the error bars in the MLS data, so this is a clear statement (we believe) regarding the model accuracy (absolute), in comparison to state-of-the art observations; we also identified a few areas of disagreement. We could add (in the revised version) percentage difference numbers regarding the "accuracy level" (model/data agreement level) for each species, if this is what the reviewer is asking. We have preferred to let the first few Figures (Figs. 1,3,4,5,6 regarding climatological levels of agreement) speak for themselves. One often obtains levels of model/data agreement within about ± 5 to 10%. However, quoting a more detailed range of "accuracies" versus species, pressure, and latitude, can add up to a fair amount of text. We have already highlighted regions where we believe that model issues might need more investigation, and some regions where data issues could also contribute to the differences (e.g., where more difficult retrievals and/or fewer data validation possibilities exist). The right panels in these climatological comparison Figures help to take into account the systematic errors in the MLS data. If the Referee really wants us to add more numbers in the text (or in the already long Abstract), we can try to do this in the revised manuscript, but we would otherwise stick to the fact that one can extract numbers out of the Figures already present in the manuscript. If another model wishes to "measure up" to the same data sets, new Figures of this kind would need to be produced, for comparison purposes.

(4) The paper also uses both the models and data to look at trends. Reading the abstract paragraph which summarises the trend work does not give me a clear view of the main scientific points that have come out of the trend work. Is there something new about the observed known recent upper stratospheric ozone increase (i.e. recovery)? Or are the main points related to whether SD or FR simulations are better for studying past trends of different types of species (and I realise there are potential issues with both approaches). The paper also discusses metrics which can be used to evaluate CCM runs using observations. There is a lot of information here but again the main messages and recommendations are not clear to me.

Reply: There are both types of aspects in our results, and while we thought that this was already fairly clear, we can try to clarify where needed, if we are given more specifics from the referee, after our revised version is finalized. Indeed, some points are made in terms of trends themselves (e.g., O_3 trends that are positive in the lower stratosphere over the MLS period, whereas longer time periods have indicated some decreases – so further confirmation with more years of data should be worthwhile in the near future), while other points clearly deal with the comparisons with model trends. In many aspects, SD-WACCM matches the latitude/height behavior of observed ozone trends quite well, and also matches the observed H_2O trends better than FR-WACCM.

For me as a reviewer the questions about this paper are

(i) what are new scientific results related to CCMs (including diagnostics) or trends in general

Reply: Please see our replies above, as this reviewer comment is mainly a summary comment.

and

(ii) why does the evaluation of the two WACCM versions belong in an ACP paper, rather than the sister journal Geophysical Model Development (GMD). At the moment, and using the abstract as a basis, I really don't get the main scientific advances which would justify ACP versus GMD.

My recommendation is that the work needs to be presented with clearer scientific messages coming through in the abstract and conclusions. Work which does not directly contribute to the ACP-level results could be put in a GMD paper, or an expanded supplement.

Reply: In response to this, we have made some changes, notably to the Abstract, main text, and conclusions, with more useful information to help strengthen the results on ozone and H_2O trends. Short of the revised version (which we are finalizing soon), please see the revised (draft) Abstract at the end of this reply, with the highlighted parts as a guide to the non-minor changes.

Stratospheric science has progressed to the point of being quite well understood from the point of view of very sophisticated tools, like SD-WACCM (with mostly correct representations, or parameterizations, of the physics and chemistry), and this limits the extent of significant new advances. However, this manuscript is (in our view) one of the more comprehensive studies that confronts such a model with multi-year and multi-species data sets, for species with different lifetimes and gradients, so that a fuller depiction of areas of agreement or disagreement can be revealed. On the trends side, there is good overall agreement within the error bars; more specifically, the degree of agreement for SD-WACCM in terms of the latitudinal and vertical patterns is actually striking (see Fig. 25 in particular), if one ignores the issue of absolute error bars. Figure 25 is also an example that could be illuminating for other model comparisons, in due time (not here). Such excellent agreement in the patterns of trends is a model success worth documenting, in our view; otherwise, it could become just "word of mouth" between modeling groups, and we feel that the actual publication is important, after careful (time-consuming) analyses. While there have been some rather broad trend comparisons in the past between averaged data sets and averaged models, there are few that go into a lot of detail for different model runs; more of this type of work may well be in preparation elsewhere.

On the issue of trimming (or splitting) the manuscript, we do feel strongly that using the Supplement is a <u>much</u> better way to help cut down somewhat on the main paper, rather than to somewhat artificially break up this comprehensive work, given that this would also require a significant amount of duplication and extra work. We believe that, after some trimming of Figures and less essential text, and other clarifications, as mentioned earlier in more detail, this paper will be improved. On the other hand, there is a need for some added text in order to explain some issues better, namely for water vapor trends, their magnitude (in relation to what one would expect from methane increases), and the differences between the two models. In the end, we feel that setting an arbitrary length goal does not make much sense, when a lot of comparison work is investigated (or even summarized) for multiple species with different lifetimes, in order to confront the models with a multi-dimensional and multi-faceted atmosphere. However, we will heed the advice regarding a trimmed down revised version, and we thank both referees for these comments.

Regarding the Journal issue, we feel quite strongly that such a paper is (or can certainly be) in the ACP domain, given that the model description is really a small part of this manuscript (WACCM having been used and described previously, including in GMD, *Morgenstern et al*., gmd-10-639-2017), and that there are some scientific results discussed here (to be further clarified, as mentioned in our replies and in the upcoming revision), even if some of this confirms past/recent work, but from our own model/data comparisons. There is some “grey region” between ACP and GMD papers, with the latter being more geared towards model description and development (if one looks through many of those articles), although there are some model evaluation papers there as well. To be more specific, we include Table 1 at the end of this reply, and this provides a summary of all the papers that are part of the current CCMI special issue, which is what we are submitting to here; this special issue encompasses several journals (including ACP and GMD). As one can see from Table 1, the more recent papers have nearly all been part of ACP, after some initial work with much more of a model description focus. Some of the articles in ACP could compare broadly to the work we are trying to present, with a combination of model and data (and comparisons). We also feel that there are detailed aspects of the MLS data sets described in our work (regarding absolute error bars and trend uncertainties, including some drift issues) that would be of much interest to the stratospheric component of the ACP readership. Without attempting to be more comprehensive, we can state that we did consider the Journal topic seriously, which also led to some delays. We also consulted with the ACP editors on this topic, and we are pleased that they agree with our views; this topic is also something that editors consider as part of the pre-review process. It is also true that going through another 4 months of review with a completely new set of reviewers and editors is a considerable burden not just on the authors, with further time delays, but also on the reviewer community (especially for longer papers). We are thus thankful for the support we obtained towards finalizing this process for ACP, and we feel that we can now focus our efforts to that end; we would very much welcome reviewer support on this aspect as well.

Minor comments

Page 1. Line 20. Can you be quantitative when discussing model over/underestimates?

Reply: Certainly, these Abstract sentences are now rewritten for clarification, as follows: "There are a few significant model/data mean biases, such as for lower stratospheric O_3, for which the models at mid- to high latitudes overestimate the mean MLS values by as much as 50% and the seasonal amplitudes by ~60%. Another clear difference occurs for HNO_3 during recurring winter periods of strong HNO_3 enhancements at high latitudes; the strong model underestimate in this case (by a factor of about 2 to 6) stems from the omission of ion chemistry relating to particle precipitation effects, in the global models used here." The relevant sections in the text will also be adjusted to match these more quantitative points.

Page 1. Lines 26-27. In what way are the detailed interactions not as well represented?

Reply: We have decided that this result, although correct, is not needed in the Abstract, given that one expects a free-running model to be less in-phase with actual dynamical situations represented better by SD-WACCM (and the observational record). This will therefore be removed from the Abstract, although the relevant (fairly brief) discussion can stay in the main text, as a demonstration of these somewhat subtle, but real differences, between model 'flavors' and observations.

Page 2. Line 12. 'differences' rather than 'variability'?

Reply: Yes, this wording is changed to 'differences'.

Page 2. Line 14. 'driven' – not the correct word for what is inside the model. Usually used for the external forcings like winds or emissions etc.

Reply: Yes, this wording is changed to "driven by time-dependent boundary conditions", without mentioning the photochemical reactions (which can be taken as a 'given', given other references to the model).

Page 2. Line 17-18. I think you should say a lot more about other SD work and cite papers, as SD v FR is a main focus of this paper. This would help to think about whether the WACCM SD results may be applicable to other SD models?

Reply: We understand the importance of comparing various SD models, and we have discussed this earlier in our reply to item (3). We plan to change part of the Introduction in the revised version to better motivate the purpose of our analyses of FR-WACCM versus SD-WACCM, as mentioned above.

Page 2. Line 27. Explain 'high quality'.

Reply: That is a fair comment, especially for a reader who might not know enough about the MLS data; however, for this Introduction, it would seem best not to try to give a detailed list of references on validation, etc… so we can just remove this somewhat vague wording for simplicity (and we are keeping the manuscript length in mind as well).

Page 2. Line 36-39. Can you give examples of trend studies that have had these problems? Again, for the trend results presented here to be of scientific interest to the community, we need to know about issues of what has been done before.

Reply: Yes, we can/will refer to some published work for ozone (Ball et al., ACP, 2017, acp-17-12269-2017) that points to regions/periods of trend differences that can be traced to data set issues and/or merging issues (for example, regarding merged SBUV data or an older ozone data version from GOZCARDS). While uncertainties relating to data merging are not easy to quantify, more work should ultimately be done on such a topic (separately from our current manuscript, of course); for SBUV, some work has been done regarding the propagation of uncertainties (*Frith et al.*, ACP, 2017, acp-17-14695-2017). Incidentally, data merging uncertainty issues point to a good reason to at least try to start using MLS data alone (as there are no data merging or sampling difference issues) for trend work, versus model results and in general.

Page 3. Line 34. After reading these sections it is not clear to me if ACE data (and which version) is included in either of the GOZCARDS versions. Please clarify.

Reply: Certainly, this text will be clarified, by changing it to: "ACE-FTS data were not included in these more recent years." The version matters less, since there is only one choice for recent years. We also plan to add the following sentence (just before paragraph 3), to clarify what was done for the special v2.20 GOZCARDS ozone product. "We note that no ACE-FTS data were included in this newer version of GOZCARDS O_3."

Page 4. Line 39. Clarify that 'organic halogens' are the source gases.

Reply: Yes, this will be changed to 'organic halogens' to specify the source.

Page 5. Line 2. So the FR WACCM is relaxed to the observed tropical winds (QBO). What is the implication of that for the comparison? Does that constrain some of the comparisons? What would happen without this relaxation? (Why is it done?).

Reply: If one wants to represent the observed stratospheric variability, one has to include QBO forcing in the tropical region; without this, the variability would be much less realistic, and less accurate. This was also the specification for the CCMI scenario (REF-C1), to include either a nudged QBO or an interactively-derived QBO (if possible). The latest version of FR-WACCM, recently released to the community, now has an interactive QBO. This was not available for this CCMI assessment.

Page 5. Line 11. New paragraph before 'Both'.

Reply: Yes, this is changed to a new paragraph.

Page 6. Line 28. The model comparisons don't use the satellite averaging kernels (or temporal sampling I suppose?). Can you add more details on why you see no reason to apply the AKs?

Reply: Some discussion of this aspect of the comparisons was already provided regarding Fig. 2 model/data differences on page 6 (lines 26-29), and this is a generic type of response for these comparisons (as has also been verified in the context of other comparisons of MLS data versus models, notably for water vapor). The MLS instrument system has sharply peaked vertical Averaging Kernels as a result of its limb viewing geometry and field of view characteristics, with stratospheric vertical resolutions of order 2.5 to 4 km in most cases (species) of relevance here. The largest impacts (from neglecting profile smoothing) can be expected in the upper troposphere, at least when comparing to fine resolution sonde profiles. Examples of smoothed and unsmoothed ozone comparisons are provided in the original MLS ozone validation paper (Fig. 6) by *Froidevaux et al.* (JGR, 2008, 10.1029/2007JD008771), in the context of comparisons versus SAGE II, which has a vertical resolution finer than 1-2 km; this shows that the effects are typically quite small (less than a few percent) even for SAGE-type profiles. The WACCM model profiles are provided on a grid that is not substantially finer than the MLS retrieval grid, and such profiles will thus be affected even less. Also, both model runs in this case are on the same vertical grid (and the model profiles do not generally differ by very large amounts); they will be affected in the same (small) way by a small amount of smoothing to match the MLS retrieval grid. While we could add more words to this effect, we will probably not plan to lengthen the manuscript much regarding this point, given that we have at least touched on this topic already.

Page 7. Line 13. Any idea why there are larger differences for SD WACCM? What are the implications for SD studies?

Reply: Transport-related model issues, as mentioned regarding some regions of disagreement between ozone observations and data, could also impact the lower stratospheric HCl abundances. However, the HCl amounts in this region are quite small, so we do not wish to over-emphasize this sort of discrepancy for this species. Finally, it is also a region where the MLS retrievals are less well constrained, in terms of percentage accuracy at least, although this does not help to alleviate model-to-model differences. We should probably not overemphasize such large percentage differences, given the low abundances in this case.

Page 7. Line 16. Explain 'good dynamical tracer' for non-experts.

Reply: Yes, we will add some words here "N_2O, a long-lived species in the lower stratosphere, which means that good (or poor) model/data agreements in this region can confirm (or deny) accurate model representations of the dynamics."

Page 12. Line 36. 'do not have the right chemistry'. I would suggest rephrasing this.

Reply: Yes, we can rephrase this to 'do not include the necessary photochemical pathways, including the effects of energetic particle precipitation on ion chemistry in the upper atmosphere'.

Revised Abstract:

We evaluate the recently delivered Community Earth System Model version 1 (CESM1) Whole Atmosphere Community Climate Model (WACCM) using satellite-derived global composition datasets, focusing on the stratosphere. The simulations include free-running (FR-WACCM) and specified-dynamics (SD-WACCM) versions of the model. Model evaluations are made using global monthly zonal mean time series obtained by the Aura Microwave Limb Sounder (MLS), as well as longer-term global data records compiled by the Global Ozone Chemistry and Related Trace gas Data Records for the Stratosphere (GOZCARDS) project. A recent update (version 2.20) to the original GOZCARDS merged ozone (O_3) data set is used here.

We discuss upper atmospheric climatology and zonal mean variability using O_3, hydrogen chloride (HCl), nitrous oxide (N_2O), nitric acid (HNO_3), and water vapor (H_2O) data. There are a few significant model/data mean biases, such as for lower stratospheric O_3, for which the models at mid- to high latitudes overestimate mean MLS values by as much as 50% and the seasonal amplitudes by ~60%; such differences require further investigations, but would appear to implicate (in part) a transport-related issue in the models. Another clear difference occurs for HNO_3 during recurring winter periods of strong HNO_3 enhancements at high latitudes; model underestimates in this case (by a factor of 2 to 6) stem from the omission of ion chemistry relating to particle precipitation effects, in the models used here. In the lower stratospheric high southern latitudes, variations in polar winter/spring composition observed by MLS are generally well matched by SD-WACCM, the main exception being for the early winter rate of decrease in HCl, which is too slow in the model. In general, we find that the latitude/pressure distributions of annual and semi-annual oscillation amplitudes derived from MLS data are properly captured by the corresponding model values.

One of the model evaluation diagnostics we use represents the closeness of fit between the model/data anomaly time series, and we also consider the correlation coefficients. Not surprisingly, SD-WACCM, which is driven by realistic dynamics, generally matches observed deseasonalized anomalies better than FR-WACCM does. We use the root mean square variability as a more valuable way to estimate differences between the two models and the observations. We find, most notably, that FR-WACCM underestimates the observed interannual variability for H_2O by ~30%, typically, and by as much as a factor of two in some regions; this has some implications for estimates of the time needed to detect small trends, based on model predictions.

We provide trend comparisons between various data sets and (CESM1) WACCM, using a multivariate linear regression (MLR) model. Both MLS and WACCM show a robust upper stratospheric O_3 increase from 2005 to 2014 by ~0.2-0.4%/yr ($\pm$ 0.2%/yr, 2σ), depending on which latitude range (tropics or mid-latitudes) is considered. In the lower stratosphere, some decreases are indicated for 1998-2014 (based on merged GOZCARDS O_3), but we find near-zero or positive trends when using MLS O_3 data alone for 2005-2014. The SD-WACCM results track these observed tendencies, although there is little statistical significance in either result; the patterns of O_3 trends versus latitude and pressure are remarkably similar between SD-WACCM and MLS. For H_2O, the most statistically significant trend for 2005-2014 is an upper stratospheric increase, peaking at slightly more than 0.5%/yr in the lower mesosphere, in fairly close agreement with SD-WACCM trends, but with smaller values in FR-WACCM. As shown before by others, there are multiple factors that can influence low-frequency variability in H_2O; indeed, these recent short-term trends are larger than what one would expect from changes associated with slow secular increases in methane. For HCl, while the lower stratospheric vertical gradients of MLS trends are duplicated to some extent by SD-WACCM, the model trends (decreases) are always on the low side of the data trends. There is also little model-based indication (in SD-WACCM) of a significantly positive HCl trend derived from the MLS tropical series at 68 hPa. These differences deserve further study. For N_2O, the MLS-derived trends (for 2005-2012) point to negative trends (of up to -1%/yr) at NH mid-latitudes and positive trends (of up to +3%/yr) at SH mid-latitudes, in good agreement with the asymmetry that exists in SD-WACCM trend results. The small observed positive N_2O trends of ~0.2%/yr in the 100 to 30 hPa tropical region are also consistent with model results (SD-WACCM in particular), which are very close to known rates of increase in tropospheric N_2O. In the case of HNO_3, MLS-derived lower stratospheric trend differences (for 2005-2014) between hemispheres are opposite in sign to those from N_2O and in reasonable agreement with both WACCM results. In general, variations tied to the QBO play a big role in terms of the interpretation of stratospheric trends over short time periods (such as 2005-2014); longer time periods are typically required to robustly extract underlying long-term trends, notably in the lower stratosphere.

The data sets and tools discussed here for the evaluation of the models could be expanded to additional comparisons of species not included here, as well as to model intercomparisons using a variety of CCMs, in order to search for systematic differences versus observations or between models, keeping in mind the range of model parameterizations and approaches.

Table 1. Pubs. in CCMI special issue (mostly ACP papers recently, with a variety of topics/thrusts).

Reference	Title	Type of study (model vs data, etc…)	Some novel aspects of atm. science?	Mostly model description or model analyses? > not much data
Jockel, P. et al. (2016), 10.5194/ **gmd**-9-1153-2016 **GMD**	Earth System Chemistry integrated Modelling (ESCiMo) with the Modular Earth Submodel System v-2.5	*One model* with different scenarios	Not really	Yes, model sensitivity (scenario) runs
Tilmes, S. et al. (2016), 10.5194/ **gmd**-9-1853-2016 **GMD**	Representation of the CESM1 CAM4-chem within the CCMI	*One model* (different scenarios) & some data	Not really	Model evaluation studies
Strode, S. A. et al. (2016), 10.5194/ **acp**-16-7285-2016	Interpreting space-based trends in CO with multiple models.	Model and data	Yes, in terms of model/data differences.	A combination of models and data
Morgenstern, O. et al. (2017), 10.5194/ **gmd**-10-639-2017 **GMD**	Review of the global models used within phase 1 of CCMI.	Descriptions of various CCMI models	Not directly	Model descriptions only
Fernandez, R. P. et al. (2017), 10.5194/**acp**-17-1673-2017	Impact of biogenic VSL bromine on the Antarctic O_3 hole during the 21st century.	*One model* and data - with model predictions	Not directly, but based on model predictions	Yes, mostly model predictions
Smalley, K. M. et al. (2017), 10.5194/ **acp**-17-8031-2017	Contribution of different processes to changes in tropical LS H_2O in CCMs.	Models and some data	Yes, based on model behaviors & inferences	Yes, mostly model analyses
Hardiman, S. C. et al. (2017), 10.5194/ **gmd**-10-1209-2017 **GMD**	The Met Office HadGEM3-ES CCM: evaluation of strat. dynamics, impact on O_3	*One model:* different simulations (FR vs SD)	Not directly	Yes, mostly model analyses and evaluations
Lin, M. et al. (2017), 10.5194/ **acp**-17-2943-2017	US surface O_3 trends & extremes (1980- 2014): quantifying the roles of rising Asian emissions, domestic controls, wildfires, and climate.	*One model* with data comparisons	Yes, based on one model's behavior & inferences	Mostly model inferences (with some data comparisons)
Maycock, A. C. et al. (2018), 10.5194/ **acp**-18-11323-2018	The representation of solar cycle signals in strat O_3- Part-2: Analysis of global models.	Mostly multi-model results	Not directly, mostly model dependence on inputs	Yes, mostly a model sensitivity study

Reference	Title	Type of study (model vs data, etc…)	Some novel aspects of atm. science?	Mostly model description or model analyses? >not much data
Morgenstern, O. et al. (2018), 10.5194/ **acp**-18-1091-2018	O_3 sensitivity to varying greenhouse gases and ozone-depleting substances in CCMI-1 simulations.	Multi-model description & consistency of responses to forcings	Not directly	Yes, a model sensitivity study
Revell, L. E. et al. (2018), 10.5194/**acp**-17-13139-2017	Impacts of Mt. Pinatubo volcanic aerosol on the tropical stratosphere in CCM simulations using CCMI & CMIP6 stratos. Aerosol data	*One model*. Sensitivity of T and O_3 response to volcanic aerosol data	Not directly	Yes, mostly a model sensitivity study
Hou, P. et al. (2018) **acp**-18-8173-2018	Sensitivity of atmos. aerosol scavenging to precip. intensity and frequency in context of climate change	Some data but mostly a prediction sensitivity study	Yes, but based on prediction sensitivities	Yes, mostly a model sensitivity study (with different met. fields)
Phalitnonkiat, P. et al. (2018), 10.5194/ **acp**-18-11927-2018	Extremal dependence between T and O_3 over the continental US.	Some data but mostly multi-model prediction	Yes, but based on model predictions	Yes, mostly a model sensitivity study
Orbe, C. et al. (2018), 10.5194/ **acp**-18-7217-2018	Large-scale tropospheric transport in the CCMI simulations.	Multi-model diffs.: AOA, transport.	Not directly	Yes, mostly a model sensitivity study
Wu, X. et al. (2018), 10.5194/ **acp**-18-7439-2018	Spatial and temporal variability of interhemispheric transport times.	*One model:* Variability of idealized tracers	To some extent, based on model sensitivity	Yes, mostly a model sensitivity study (of variability)
Dietmuller, S. et al. (2018), 10.5194/ **acp**-18-6699-2018	Quantifying the effect of mixing on the mean age of air in CCMVal-2 and CCMI-1 models.	Multi-model look: factors influencing AOA	Not directly	Yes, mostly a model sensitivity study
Dhomse, S. S. et al. (2018), 10.5194/ **acp**-18-8409-2018	Estimates of ozone return dates from CCMI simulations.	Multi-model estimates: O_3 return dates	Yes, but based on predictions	Yes, mostly a model sensitivity study
Ayarzaguena, B. et al. (2018), 10.5194/ **acp**-18-11277-2018	No robust evidence of future changes in major stratospheric sudden warmings: a multi-model CCMI assessment	Multi-model study of major strat. sudden warmings	Yes, based on model predictions	Yes, mostly a model sensitivity study

Reference	Title	Type of study (model vs data, etc…)	Some novel aspects of atm. science?	Mostly model description or model analyses? >not much data
Lamy, K. et al. (2018), ACPD, 10.5194/**acp**-2018-525	UV radiation modelling using output from the CCMI	Multi-model UVI versus climo UVI data	Yes, based on model results	A combination of models and data
Revell, L. E. et al. (2018) **acp**-2018-615	Tropospheric ozone in CCMI models and Gaussian emulation to understand biases in the SOCOLv3 CCM.	Multi-model comparison of tropos. ozone vs data	Mostly geared towards model refinements	A combination of models and data

Reply to the review from Referee 2

We are thankful to this referee for the review and the associated suggestions, listed in italics below. We provide our detailed responses (regular font) and plans; our revised manuscript will be available in a fairly short time. For added information, we have provided the revised Abstract in our reply here, although most of those (highlighted) changes were done as a response to comments from the other Referee.

It would seem from the referee comments that there are no demonstrable big issues with the science (or math), besides some requested clarifications, and we are pleased that Referee 2 found our manuscript to contain "a lot of valuable and detailed information" [and Referee 1 also found "some interesting results"].

The manuscript aims to evaluate the stratospheric composition of the free-running and specified-dynamics version of CESM1 (WACCM). The evaluations are based on comparisons to satellite measurements including single-instrument and merged data records. The model diagnostics include zonal monthly mean comparisons, seasonal and semi-annual cycles as well as long-term trends. All evaluations are described in detail and valuable information on various aspects of the model performance is provided. Overall, the manuscript is of great interest for scientist directly working with WACCM or potentially with other earth-system models. Therefore, such a detailed manuscript would seem much more appropriate in a journal focused on geoscientific model development/validation and I would urge the authors to submit it to a journal focused on this topic.

Reply: We do not agree that the fairly comprehensive analyses presented here are of more limited interest to modeling groups only, because there are also inferences from data sets that have not been presented before (in particular trend analyses from MLS data alone). We have also added a few clarifications to better explain certain aspects of these trends and comparisons (largely in response to the other Referee).

Regarding the Journal issue, we feel quite strongly that such a paper is (or can certainly be) in the ACP domain, given that the model description is really a small part of this manuscript (WACCM having been used and described previously, including in GMD, *Morgenstern et al.*, gmd-10-639-2017), and that there are some scientific results discussed here (to be further clarified, as mentioned in our replies and in the upcoming revision), even if some of this confirms past/recent work, but from our own model/data comparisons. There is some "grey region" between ACP and

GMD papers, with the latter being more geared towards model description and development (if one looks through many of those articles), although there are some model evaluation papers there as well. To be more specific, we include Table 1 at the end of this reply, and this provides a summary of all the papers that are part of the current CCMI special issue, which is what we are submitting to here; this special issue encompasses several journals (including ACP and GMD). As one can see from Table 1, the more recent papers have nearly all been part of ACP, after some initial work with much more of a model description focus. Some of the articles in ACP could compare broadly to the work we are trying to present, with a combination of model and data (and comparisons). We also feel that there are detailed aspects of the MLS data sets described in our work (regarding absolute error bars and trend uncertainties, including some drift issues) that would be of much interest to the stratospheric component of the ACP readership. Without attempting to be more comprehensive, we can state that we did consider the Journal topic seriously, which also led to some delays. We also consulted with the ACP editors on this topic, and we are pleased that they agree with our views; this topic is also something that editors consider as part of the pre-review process. It is also true that going through another 4 months of review with a completely new set of reviewers and editors is a considerable burden not just on the authors, with further time delays, but also on the reviewer community (especially for longer papers). We are thus thankful for the support we obtained towards finalizing this process for ACP, and we feel that we can now focus our efforts to that end; we would very much welcome reviewer support on this aspect as well.

Major comments

1) The paper delivers a lot of valuable and detailed information, however, is overall very long. In particular, the number of figures could be reduced from 32 to around 20. To give one example, Figure 2 is only discussed very briefly in the text in order to illustrate mean biases and annual cycle differences shown elsewhere and could be removed.

Reply: We really prefer to keep Fig. 2 in the main text as this does show much better than Fig. 1 how the models and the data differ (in certain regions) in terms of not only the mean differences, but also the annual cycle; these differences are now also better quantified in the upcoming revised version (in answer to a comment from reviewer 1).

We are planning to cut down on the length of this manuscript, mainly by relegating some of the less critical Figures to the Supplement. Although this does not necessarily translate into a very large cut in terms of text length, we consider this work to be a fairly comprehensive analysis, which therefore leads to a longer paper; there have definitely been some longer (atmospheric) papers in the literature, and specifically in ACP. Turning this into two separate papers mainly for the sake of overall length seems too artificial, and this would be quite an elaborate proposition, with the need for some duplication regarding both the data sets and the models; as an aside, this would actually lead to more reviewing work for the community. We hope to have shown that detailed analyses are necessary to enable identification of both good agreement (a result in itself) or significant differences between model runs and the data sets, but also for some of the more subtle differences, and furthermore, that an understanding and discussion of error bars and potential data issues is important. We will also strive to reduce the amount of text in the revised manuscript, especially where some less critical aspects can be discussed more succinctly, or taken out altogether. In particular, we plan to shorten Section 5.1.1 (pages 11-13) to a text length that roughly matches (rather than exceeds) the text length of Section 5.1.2 (on variability issues); the cuts to Sect. 5.1.1 will be of order 30% (or more).

In terms of reducing the amount of Figures and related changes, our specific plans are to remove Figs. 13, 14, 15, and 17 from the main text (and relegate these to the Supplement, with a slightly shortened discussion), since these mainly reinforce the expectation (already noted for O_3 and H_2O) of better model/data fits from SD-WACCM, as one might expect from a model with better dynamical constraints than the FR-WACCM version. Such an expectation does not hold for the variability diagnostics, so these are really best left in the main text, although we will plan to displace Figure 22 (on the N_2O and HNO_3 variability comparisons), and move it to the Supplement. Moreover, we feel that Figure 31 on lat/p contours of short-term trend for various species can be moved to the Supplement, as it is less critical, and given past (and ongoing) work on this topic. While Figure 32 is interesting to us, it is more of a side note on lower stratospheric tropical cohesiveness for various species exhibiting similar dynamical variability, so we decided that the text and Figure in this case can be eliminated altogether without much of an impact on this paper.

In summary, the total number of Figures in the main text will be trimmed down by almost a quarter, with a more manageable total of 25 Figures; writing up a multi-year effort of (part-time) work on detailed model/data comparisons is bound to lead to a longer manuscript than several shorter analyses; to our knowledge, fits, correlations, variability, and trend comparisons are rarely investigated to this extent in model/data comparisons, even for a single model (or two flavors of one model). This, with some reductions (and clarifications) in the text (including the Abstract and Conclusions section), will at least show our good faith effort towards the referee comments. Recommending a goal of exactly 20 Figs. is rather arbitrary, but our point here is that we have considered these requests with some care, and that we are being responsive.

2) Differences are often only listed and not explored more in detail. To give one example, model HCl shows systematic differences in the lower stratosphere (evaluation based on Fig. 4) and a discussion relating those differences to shortcomings in the model transport or model chemistry would be interesting. Given the length of the manuscript, one could focus on the gases for which the detected differences are discussed in terms of model behavior (e.g., HNO_3). Differences for other gases can be mentioned in the manuscript with the according figures being moved to the supplement.

Reply: Yes, we pursued this type of reorganization, as explained above, with what we would consider a reasonable amount of delegating of material to the Supplement. We find some value in the remaining Figures, and feel that using a somewhat arbitrary number (such as 20) is not justified for a paper that covers a fair amount of ground and wishes to confront the models with a multi-species approach, in order to check for potential areas of weakness. Just stating good agreement and putting almost every Figure in the Supplement could work also, in principle, but that would be the other extreme, with a nearly complete lack of visual confirmation, which we think is important to preserve as well. Also, while we are striving to cut down on the length here, there are other long papers in the literature (but we will most likely avoid this sort of length in the future).

3) In section 3, existing evaluations of WACCM and the WACCM composition in particular should be discussed. Such references come up in the latter part of the manuscript. If they are given combined in this section, it will easier for the reader to identify what the current challenges are and what is new in this manuscript.

Reply: Yes, this section and/or the Introduction will be modified in the revised version (without adding too much length) to take this into account; in particular, we will add some motivation for the comparisons done here for FR-WACCM versus SD-WACCM (and observations).

Minor comments:

1) Consider changing the title to 'Evaluation of CESM1 (WACCM) free-running and specified-dynamics stratospheric composition simulations using global multi-species satellite data records.

Reply: Given that water vapor is considered all the way through the mesosphere, we prefer to stick to our original title, but we did consider this suggestion.

2) Page 5, line 31 – Page 6, line 2: This text could be moved to the discussion of the MLS data record in section 2.1.

Reply: While this could be done in principle, we feel that the species-specific discussions of error bars and validation work is really best kept as part of the discussions for each species, and that the flow is less awkward this way; we have thus not tried to reorganize these portions of text.

3) Page 7, line 24: Do you mean all earth system model or just WACCM with the term 'general model underestimation'?

Reply: We mean just the WACCM models here. This is clarified in the revised version by stating "model underestimation by both WACCM versions." However, it is implicit that other models without the proper (more complicated) chemical processes and energetic particle pathways will also underestimate HNO_3 in the same fashion.

4) Page 9, line 7 -10: Here, and also in other places, the sentence is too long for easy understandability. Consider splitting into two sentences at the semicolon.

Reply: Yes, we will start a new sentence instead of using a semi-colon, if/as that may help. We will also consider some other places for such an issue.

5) Page 12, line 5-8: The statement is made for the upper mesosphere. But isn't it also true for the stratosphere?

Reply: The statement (regarding worse diagnostic values) is somewhat true for the upper stratosphere as well, but we are mainly referring to SD-WACCM here; nevertheless, we have modified the revised text to state that the (SD-WACCM) diagnostics “are of poorest quality in the mesosphere” (etc…).

6) Page 13, line 7: MIPAS has been used earlier in the manuscript.

Reply: Yes, thank you; this is readily fixed by defining MIPAS earlier on in the text (in the 2nd part of section 4).

Revised Abstract:

We evaluate the recently delivered Community Earth System Model version 1 (CESM1) Whole Atmosphere Community Climate Model (WACCM) using satellite-derived global composition datasets, focusing on the stratosphere. The simulations include free-running (FR-WACCM) and specified-dynamics (SD-WACCM) versions of the model. Model evaluations are made using global monthly zonal mean time series obtained by the Aura Microwave Limb Sounder (MLS), as well as longer-term global data records compiled by the Global Ozone Chemistry and Related Trace gas Data Records for the Stratosphere (GOZCARDS) project. A recent update (version 2.20) to the original GOZCARDS merged ozone (O_3) data set is used here.

We discuss upper atmospheric climatology and zonal mean variability using O_3, hydrogen chloride (HCl), nitrous oxide (N_2O), nitric acid (HNO_3), and water vapor (H_2O) data. There are a few significant model/data mean biases, such as for lower stratospheric O_3, for which the models at mid- to high latitudes overestimate mean MLS values by as much as 50% and the seasonal amplitudes by ~60%; such differences require further investigations, but would appear to implicate (in part) a transport-related issue in the models. Another clear difference occurs for HNO_3 during recurring winter periods of strong HNO_3 enhancements at high latitudes; model underestimates in this case (by a factor of 2 to 6) stem from the omission of ion chemistry relating to particle precipitation effects, in the models used here. In the lower stratospheric high southern latitudes, variations in polar winter/spring composition observed by MLS are generally well matched by SD-WACCM, the main exception being for the early winter rate of decrease in HCl, which is too slow in the model. In general, we find that the latitude/pressure distributions of annual and semi-annual oscillation amplitudes derived from MLS data are properly captured by the corresponding model values.

One of the model evaluation diagnostics we use represents the closeness of fit between the model/data anomaly time series, and we also consider the correlation coefficients. Not surprisingly, SD-WACCM, which is driven by realistic dynamics, generally matches observed deseasonalized anomalies better than FR-WACCM does. We use the root mean square variability as a more valuable way to estimate differences between the two models and the observations. We find, most notably, that FR-WACCM underestimates the observed interannual variability for H_2O by ~30%, typically, and by as much as a factor of two in some regions; this has some implications for estimates of the time needed to detect small trends, based on model predictions.

We provide trend comparisons between various data sets and (CESM1) WACCM, using a multivariate linear regression (MLR) model. Both MLS and WACCM show a robust upper stratospheric O_3 increase from 2005 to 2014 by ~0.2-0.4%/yr (± 0.2%/yr, 2σ), depending on which latitude range (tropics or mid-latitudes) is considered. In the lower stratosphere, some decreases are indicated for 1998-2014 (based on merged GOZCARDS O_3), but we find near-zero or positive trends when using MLS O_3 data alone for 2005-2014. The SD-WACCM results track these observed tendencies, although there is little statistical significance in either result; the patterns of O_3 trends versus latitude and pressure are remarkably similar between SD-WACCM and MLS. For H_2O, the most statistically significant trend for 2005-2014 is an upper stratospheric increase, peaking at slightly more than 0.5%/yr in the lower mesosphere, in fairly close agreement with SD-WACCM trends, but with smaller values in FR-WACCM. As shown before by others, there are multiple factors that can influence low-frequency variability in H_2O; indeed, these recent short-term trends are larger than what one would expect from changes associated with slow secular increases in methane. For HCl, while the lower stratospheric vertical gradients of MLS trends are duplicated to some extent by SD-WACCM, the model trends (decreases) are always on the low side of the data trends. There is also little model-based indication (in SD-WACCM) of a significantly positive HCl trend derived from

the MLS tropical series at 68 hPa. These differences deserve further study. For N_2O, the MLS-derived trends (for 2005-2012) point to negative trends (of up to -1%/yr) at NH mid-latitudes and positive trends (of up to +3%/yr) at SH mid-latitudes, in good agreement with the asymmetry that exists in SD-WACCM trend results. The small observed positive N_2O trends of ~0.2%/yr in the 100 to 30 hPa tropical region are also consistent with model results (SD-WACCM in particular), which are very close to known rates of increase in tropospheric N_2O. In the case of HNO_3, MLS-derived lower stratospheric trend differences (for 2005-2014) between hemispheres are opposite in sign to those from N_2O and in reasonable agreement with both WACCM results. In general, variations tied to the QBO play a big role in terms of the interpretation of stratospheric trends over short time periods (such as 2005-2014); longer time periods are typically required to robustly extract underlying long-term trends, notably in the lower stratosphere.

The data sets and tools discussed here for the evaluation of the models could be expanded to additional comparisons of species not included here, as well as to model intercomparisons using a variety of CCMs, in order to search for systematic differences versus observations or between models, keeping in mind the range of model parameterizations and approaches.

Table 1. Pubs. in CCMI special issue (mostly ACP papers recently, with a variety of topics/thrusts).

Reference	Title	Type of study (model vs data, etc…)	Some novel aspects of atm. science?	Mostly model description or model analyses? > not much data
Jockel, P. et al. (2016), 10.5194/ **gmd**-9-1153-2016 **GMD**	Earth System Chemistry integrated Modelling (ESCiMo) with the Modular Earth Submodel System v-2.5	*One model* with different scenarios	Not really	Yes, model sensitivity (scenario) runs
Tilmes, S. et al. (2016), 10.5194/ **gmd**-9-1853-2016 **GMD**	Representation of the CESM1 CAM4-chem within the CCMI	*One model* (different scenarios) & some data	Not really	Model evaluation studies
Strode, S. A. et al. (2016), 10.5194/ **acp**-16-7285-2016	Interpreting space-based trends in CO with multiple models.	Model and data	Yes, in terms of model/data differences.	A combination of models and data
Morgenstern, O. et al. (2017), 10.5194/ **gmd**-10-639-2017 **GMD**	Review of the global models used within phase 1 of CCMI.	Descriptions of various CCMI models	Not directly	Model descriptions only
Fernandez, R. P. et al. (2017), 10.5194/**acp**-17-1673-2017	Impact of biogenic VSL bromine on the Antarctic O_3 hole during the 21st century.	*One model* and data - with model predictions	Not directly, but based on model predictions	Yes, mostly model predictions
Smalley, K. M. et al. (2017), 10.5194/ **acp**-17-8031-2017	Contribution of different processes to changes in tropical LS H_2O in CCMs.	Models and some data	Yes, based on model behaviors & inferences	Yes, mostly model analyses
Hardiman, S. C. et al. (2017), 10.5194/ **gmd**-10-1209-2017 **GMD**	The Met Office HadGEM3-ES CCM: evaluation of strat. dynamics, impact on O_3	*One model:* different simulations (FR vs SD)	Not directly	Yes, mostly model analyses and evaluations
Lin, M. et al. (2017), 10.5194/ **acp**-17-2943-2017	US surface O_3 trends & extremes (1980- 2014): quantifying the roles of rising Asian emissions, domestic controls, wildfires, and climate.	*One model* with data comparisons	Yes, based on one model's behavior & inferences	Mostly model inferences (with some data comparisons)
Maycock, A. C. et al. (2018), 10.5194/ **acp**-18-11323-2018	The representation of solar cycle signals in strat O_3- Part-2: Analysis of global models.	Mostly multi-model results	Not directly, mostly model dependence on inputs	Yes, mostly a model sensitivity study

Reference	Title	Type of study (model vs data, etc…)	Some novel aspects of atm. science?	Mostly model description or model analyses? >not much data
Morgenstern, O. et al. (2018), 10.5194/ **acp**-18-1091-2018	O_3 sensitivity to varying greenhouse gases and ozone-depleting substances in CCMI-1 simulations.	Multi-model description & consistency of responses to forcings	Not directly	Yes, a model sensitivity study
Revell, L. E. et al. (2018), 10.5194/**acp**-17-13139-2017	Impacts of Mt. Pinatubo volcanic aerosol on the tropical stratosphere in CCM simulations using CCMI & CMIP6 stratos. Aerosol data	*One model*. Sensitivity of T and O_3 response to volcanic aerosol data	Not directly	Yes, mostly a model sensitivity study
Hou, P. et al. (2018) **acp**-18-8173-2018	Sensitivity of atmos. aerosol scavenging to precip. intensity and frequency in context of climate change	Some data but mostly a prediction sensitivity study	Yes, but based on prediction sensitivities	Yes, mostly a model sensitivity study (with different met. fields)
Phalitnonkiat, P. et al. (2018), 10.5194/ **acp**-18-11927-2018	Extremal dependence between T and O_3 over the continental US.	Some data but mostly multi-model prediction	Yes, but based on model predictions	Yes, mostly a model sensitivity study
Orbe, C. et al. (2018), 10.5194/ **acp**-18-7217-2018	Large-scale tropospheric transport in the CCMI simulations.	Multi-model diffs.: AOA, transport.	Not directly	Yes, mostly a model sensitivity study
Wu, X. et al. (2018), 10.5194/ **acp**-18-7439-2018	Spatial and temporal variability of interhemispheric transport times.	*One model:* Variability of idealized tracers	To some extent, based on model sensitivity	Yes, mostly a model sensitivity study (of variability)
Dietmuller, S. et al. (2018), 10.5194/ **acp**-18-6699-2018	Quantifying the effect of mixing on the mean age of air in CCMVal-2 and CCMI-1 models.	Multi-model look: factors influencing AOA	Not directly	Yes, mostly a model sensitivity study
Dhomse, S. S. et al. (2018), 10.5194/ **acp**-18-8409-2018	Estimates of ozone return dates from CCMI simulations.	Multi-model estimates: O_3 return dates	Yes, but based on predictions	Yes, mostly a model sensitivity study
Ayarzaguena, B. et al. (2018), 10.5194/ **acp**-18-11277-2018	No robust evidence of future changes in major stratospheric sudden warmings: a multi-model CCMI assessment	Multi-model study of major strat. sudden warmings	Yes, based on model predictions	Yes, mostly a model sensitivity study

Reference	**Title**	**Type of study (model vs data, etc…)**	**Some novel aspects of atm. science?**	**Mostly model description or model analyses? >not much data**
Lamy, K. et al. (2018), ACPD, 10.5194/**acp**-2018-525	UV radiation modelling using output from the CCMI	Multi-model UVI versus climo UVI data	Yes, based on model results	A combination of models and data
Revell, L. E. et al. (2018) **acp**-2018-615	Tropospheric ozone in CCMI models and Gaussian emulation to understand biases in the SOCOLv3 CCM.	Multi-model comparison of tropos. ozone vs data	Mostly geared towards model refinements	A combination of models and data

Evaluation of CESM1 (WACCM) free-running and specified-dynamics atmospheric composition simulations using global multi-species satellite data records

Lucien Froidevaux[1], Douglas E. Kinnison[2], Ray Wang[3], John Anderson[4], and Ryan A. Fuller[1]

[1] Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA
[2] National Center for Atmospheric Research, Boulder, CO, USA
[3] School of Earth and Atmospheric Sciences, Georgia Institute of Technology, Atlanta, GA, USA
[4] School of Science, Hampton University, Hampton, VA, USA

Correspondence to: L. Froidevaux (lucienf@jpl.nasa.gov)

Abstract

We evaluate the recently delivered Community Earth System Model version 1 (CESM1) Whole Atmosphere Community Climate Model (WACCM) using satellite-derived global composition datasets, focusing on the stratosphere. The simulations include free-running (FR-WACCM) and specified-dynamics (SD-WACCM) versions of the model. Model evaluations are made using global monthly zonal mean time series obtained by the Aura Microwave Limb Sounder (MLS), as well as longer-term global data records compiled by the Global Ozone Chemistry and Related Trace gas Data Records for the Stratosphere (GOZCARDS) project. A recent update (version 2.20) to the original GOZCARDS merged ozone (O_3) data set is used here.

We discuss upper atmospheric climatology and zonal mean variability using O_3, hydrogen chloride (HCl), nitrous oxide (N_2O), nitric acid (HNO_3), and water vapor (H_2O) data. There are a few significant model/data mean biases, such as for lower stratospheric O_3, for which the models at mid- to high latitudes overestimate ~~the~~ mean MLS~~observed~~ values by as much as 50% and the seasonal amplitudes by ~60%; such differences require further investigations, but would appear to implicate (in part) a transport-related issue in the models. ~~-~~Another clear difference occurs for HNO_3 during recurring winter periods of strong HNO_3 enhancements at high latitudes; ~~th~~model underestimates in this case (by a factor of 2 to 6) ~~is~~ stem~~s~~ from the ~~known~~ omission of ion chemistry relating to particle precipitation effects, in the ~~global~~ models used here. In the lower stratospheric~~e at~~ high southern latitudes, ~~the~~ variations in polar winter/spring composition observed by MLS are generally well matched by SD-WACCM, the main exception being for the early winter rate of decrease in HCl, which is too slow in the model. In general, we find that the latitude/pressure distributions of annual and semi-annual oscillation amplitudes derived from ~~the~~ MLS data are properly captured by the corresponding model values. ~~Nevertheless, detailed aspects of the interactions between the quasi-biennial, annual, and semi-annual ozone variations in the upper stratosphere are not as well represented by FR-WACCM as by SD-WACCM.~~

One of the model evaluation diagnostics we use represents the closeness of fit between the model/data anomaly time series, and we also consider the correlation coefficients. Not surprisingly, SD-WACCM, which is driven by realistic dynamics, generally matches observed deseasonalized anomalies better than FR-WACCM does. ~~Other results indicate that the root mean square variability~~ We use the root mean square variability as a more valuable way to estimate differences between the two models and the observations. ~~is sometimes found to be significantly smaller in FR-WACCM than in SD-WACCM and the observations. Most notably,~~ We find, most notably, that FR-WACCM underestimates the observed interannual variability for H_2O by ~30%, typically, and by as much as a factor of two in some regions; this has some implications for estimates of the time needed to detect small trends, based on model predictions.

We provide trend comparisons between various data sets and (CESM1) WACCM, ~~have derived trends~~ using a multivariate linear regression (MLR) model. ~~, and there is a robust signal in B~~both MLS ~~observations~~ and WACCM show a robust ~~of an~~ upper stratospheric O_3 increase from 2005 to 2014 by ~0.2-0.4%/yr (± 0.2%/yr, 2σ), depending on which ~~broad~~ latitude range~~bin~~ (tropics or mid-latitudes) is considered. In the lower stratosphere, ~~while~~ some decreases are indicated for 1998-2014 (based on merged GOZCARDS O_3), but ~~,~~ we find near-zero or positive trends when using MLS O_3 data alone for 2005-2014. ~~, albeit with no robust statistical significance.~~ The SD-WACCM results track these observed ~~such positive~~ tendencies, ~~(al~~though~~beit with no~~ there is little statistical ~~statistical~~ significance in either result; ~~).~~ the patterns of O_3 trends versus latitude and pressure are remarkably similar between SD-WACCM and MLS. For H_2O, the most statistically significant ~~trend~~ trend~~result~~ for 2005-2014 is

an upper stratospheric increase, peaking at slightly more than 0.5%/yr in the lower mesosphere, in fairly close agreement with SD-WACCM trends, but with smaller values in FR-WACCM. As shown before by others, there are multiple factors that can influence low-frequency variability in H_2O; indeed, these recent short-term trends are larger than what one would expect from changes associated with slow secular increases in methane. For HCl, while the lower stratospheric vertical gradients of MLS trends are duplicated to some extent by SD-WACCM, the model trends (decreases) are always on the low side of the data trends. There is also little model-based indication (in SD-WACCM) of a significantly positive HCl trend derived from the MLS tropical series at 68 hPa.~~;~~ T~~t~~hese~~is~~ differences deserve~~s~~ further study. For N_2O, the MLS-derived trends (for 2005-2012) point to negative trends (of up to ~~about~~ -1%/yr) at~~in the~~ NH mid-latitudes and positive trends (of up to ~~about~~ +3%/yr) at~~in the~~ SH mid-latitudes, in good agreement with the asymmetry that exists in SD-WACCM trend results. The s~~The s~~mall observed positive N_2O ~~N_2O~~ trends of ~0.2%/yr in the 100 to 30 hPa tropical region are also consistent with model results (SD-WACCM in particular), which ~~in turn~~ are very close to ~~the~~ known rates of increase in tropospheric N_2O. In the case of HNO_3, MLS-derived lower stratospheric trend differences (for 2005-2014) between hemispheres are opposite in sign to those from N_2O and in reasonable agreement with both WACCM results~~, despite large error bars~~. In general, variations tied to the QBO play a big role in terms of the interpretation of stratospheric trends over short time periods (such as 2005-2014); longer time periods are typically required to robustly extract underlying long-term trends, notably in the lower stratosphere.

The data sets and tools discussed here for the evaluation of the models could be expanded to additional comparisons of species not included here, as well as to model intercomparisons using a variety of CCMs, in order to search for systematic differences versus observations or between models, keeping in mind the range of model ~~at there are different~~ parameterizations and approaches. ~~for both free-running and specified-dynamics simulations.~~

1 Introduction

State-of-the art chemistry climate models (CCMs) are known to reproduce the main features of stratospheric climatology and change, although there have~~s~~ always been some differences~~variability~~ between ~~the~~ models (e.g., Waugh and Eyring, 2008; SPARC, 2010; Dhomse et al., 2018). Free-running CCMs are used to make long-term simulations of atmospheric composition, as well as predictions of future changes, driven by ~~a large set of photochemical reactions, as well as~~ time-dependent boundary conditions for surface concentrations of greenhouse gases and ozone depleting substances (ODSs), sea surface temperatures and sea ice concentrations, 11-year solar variability, sulfate aerosol surface area density, as well as tropospheric ozone and aerosol precursor emissions. In more recent years, modeling groups have implemented "specified-dynamics" versions that are constrained to meteorological fields (e.g., surface pressure, temperature, and ~~horizontal and meridional~~ winds). Our main purpose ~~here~~ is to evaluate these two types of model runs from CESM1 WACCM, using multi-species satellite-derived global composition data sets; we ~~will~~ refer to these two types as FR-WACCM (~~, for the~~ free-running model) ~~version~~, and SD-WACCM~~,~~ (~~for the~~ specified dynamics version). The SD-WACCM version has been used in studies ranging from examination of ozone trends (e.g., Solomon et al., 2016; Ball et al., 2017; Wilka et al., 2018) to evaluation of galactic cosmic ray influence on ozone (Jackman et al., 2016). This configuration has also been used to study dynamical processes that affect stratospheric ozone (e.g., Khosrawi et al., 2013; Gille et al., 2014) and has contributed to the understanding of satellite occultation instrument differences (Sakazaki et al., 2015). Here, we perform a detailed model/observational analysis of two configurations (FR and SD) using the same modeling system (CESM), and identical tracer advection and chemistry modules. Differences between the two configurations should be caused mainly by the influence of different temperature fields on chemistry and by different mean circulations. The ~~model~~ simulations a~~we~~re based on scenarios defined by the Chemistry Climate Model Initiative (CCMI) (Eyring et al., 2013; Morgenstern et al., 2017). Our evaluation focus is on monthly zonal mean ~~time~~ series from the models versus satellite-derived ~~global~~ data~~ sets~~. The main stratospheric ~~(and mesospheric) time~~ series used here are from Aura MLS ~~products~~ data (version 4.2~~ data~~) and from the Global OZone Chemistry And Related trace gas Data records for the Stratosphere (GOZCARDS), which include MLS data from late 2004 onward. ~~The~~ GOZCARDS~~ data records~~ includes merged multi-satellite data files for O_3, H_2O, and HCl, and Aura MLS~~-derived~~ data~~ files~~ for HNO_3 and N_2O; these 5 species are used for the ~~model~~

evaluations~~ herein~~. We ~~also ~~focus, in part, on the ~~Aura ~~MLS ~~high quality ~~data sets for 2005-2014 (~~with ~~2014 being the last year of WACCM runs considered here). The regular and nearly uninterrupted daily global coverage of the MLS day and night measurements leads to minimal sampling-related biases, both for climatological comparisons (~~see ~~Toohey et al., 2013) and trend-related studies. This data set also has a well characterized set of error bars (see Livesey et al., 2018, for the latest update to the data quality documentation); however, we also note that there are some caveats to take into account regarding long-term stability for some of the MLS species.

In terms of ~~the ~~model/data comparisons, we ~~will ~~analyze the climatological mean state and "goodness of fit" issues, as well as variability. While one has the expectation that, in general, better fits to the data would be obtained for a specified-dynamics run than for a less dynamically constrained run (FR-WACCM), one needs to demonstrate this ~~quantitatively ~~with diagnostics that ~~can~~ provide enough differentiation between models that ~~are ~~can~~sometimes~~ ~~found to ~~track each other closely. There have been essentially no ~~published ~~trend studies using ~~the ~~Aura MLS data ~~set ~~by itself. T~~However, t~~his data set now covers a sufficiently long ~~time ~~period that it becomes useful to investigate such trends, as the analyses deal with one ~~homogeneous ~~data set only, wh~~ile~~ich can remove~~ing the~~ potential issues ~~associated ~~with data merging prior to 2005, whether related to poorer sampling ~~,~~ or to uncertainties in ~~the ~~bias removal between ~~various ~~data sets. While such uncertainties can be difficult to quantify, attempts have been made in the case of ozone data merging from multiple SBUV instruments, which display relative biases and drifts (Frith et al., 2017); data merging uncertainties in this case were shown to play a large role regarding overall trend uncertainties. Regarding sampling issues, Millán et al. (2016) showed, based on simulated atmospheric fields, that solar occultation-type sampling can significantly bias trend results, as well as increase the time period required for robust trend detection, compared to emission-type (much denser) sampling. On the downside, a shorter time series will ~~also ~~lead to larger uncertainties in ~~the ~~derived trends.

We ~~first ~~provide (in Sect. 2) an overview of the ~~the ~~global stratospheric data sets used here~~for these comparisons~~. Brief descriptions of FR-WACCM and SD-WACCM are given ~~provided~~ in Sect. 3. Climatological comparisons between ~~the ~~models and ~~the ~~Aura MLS data~~ sets~~ are provided in Sect. 4, in order to assess~~, for example,~~ whether any obvious biases exist; th~~ese~~is ~~comparisons~~ includes an overview of the main short-term variations, namely the annual oscillation (AO) and semi-annual oscillation (SAO). More detailed comparisons of deseasonalized anomaly time series are provided in Sect. 5.1, where we evaluate how well the two model versions fit the data sets, both in terms of closeness of fits and variability. T~~Finally, t~~rend results and comparisons are provid~~investigat~~ed in Sect. 5.2, before the closing summary and discussion in Sect. 6.~~.~~

2 Data sets

2.1 Aura MLS

The Microwave Limb Sounder (MLS) is one of four instruments on NASA's Aura satellite, launched on July 15, 2004. The MLS antenna scans the atmospheric limb as Aura orbits the Earth in a near-polar sun-synchronous orbit, and the instrument measures thermal emission (day and night) in narrow spectral channels, via microwave radiometers operating at frequencies near 118, 190, 240, and 640 GHz, as well as a 2.5 THz module to measure OH. MLS (see Waters et al., 2006) has been providing a variety of daily vertical stratospheric temperature and composition profiles (~3500 profiles per day per product), with some measurements extending down to the upper tropospheric region, and some into the upper mesosphere or higher. For more information and access to the MLS data, the reader is referred to http://disc.sci.gsfc.nasa.gov/Aura/data-holdings/MLS; the current data version is labeled 4.2*x* (with *x* varying between 0 and 3, depending on the date). Data users interested in MLS data quality and characterization, estimated errors, and related information, should consult Livesey et al. (2018), the latest update to the MLS data quality document (available from the MLS website at http://mls.jpl.nasa.gov~~-~~).

2.2 GOZCARDS

The data set considered here for longer-term model evaluation analyses is from GOZCARDS, a data record ~~that was~~ created using ~~high quality~~ satellite-based Level 2 data as "source" data sets, which were merged ~~together~~ into global monthly zonal mean ~~record~~s for O_3, H_2O, and HCl, going back in time before the (2004) launch of Aura. Readers are referred to the GOZCARDS description and highlights provided by Froidevaux et al. (2015). In brief, for O_3, the ~~original~~ GOZCARDS version 1.01 (v1.01) data record starts in 1979 with solar occultation data~~measurements~~ from the first Stratospheric Aerosol and Gas Experiment (SAGE I), and continues with data from SAGE II, the Halogen Occultation Experiment (HALOE), the Upper Atmosphere Research Satellite (UARS) MLS, the Atmospheric Chemistry Experiment Fourier Transform Spectrometer (ACE-FTS, using solar occultation), and Aura MLS. Basically, the overlap time periods for different data sets are used to calculate offsets between zonal mean time series in 10°-wide latitude bins at each pressure level, and the data sets are adjusted to a reference value (SAGE II mean values, for O_3, or an average of satellite measurements for H_2O and HCl).~~;~~ M~~m~~onthly standard deviations are also provided, along with other diagnostic quantities. GOZCARDS data extensions past 2012 were created simply by adding more recent MLS data, appropriately adjusted to account for zonal mean differences between versions, once the MLS O_3 v2.2 data became unavailable (for 2013 onward); ~~the latest~~ ACE-FTS data ~~version~~ were~~as~~ not included in these more recent years. GOZCARDS O_3 ~~(v1.01)~~ has been used for past O_3 trend assessments and in comparisons to other data records (e.g., WMO, 2014; Nair et al., 2015; Tummon et al., 2015; Harris et al., 2015; Ball et al., 2017, 2018).

For the GOZCARDS ~~ozone~~ O_3 data discussed here, unless otherwise noted, we use GOZCARDS v2.20, a recent improvement and update to the original version. Th~~e ozone v2.20~~is data set was provided for an updated assessment of stratospheric ozone by Steinbrecht et al. (2017), as well as for the assessment activities of the Long-term Ozone Trends and Uncertainties in the Stratosphere (LOTUS) project and in preparation for the latest international report on the state of the ozone layer, led by the World Meteorological Organization (WMO). GOZCARDS ozone data updates are also used as part of the yearly "State of the Climate" stratospheric ozone-related summaries, produced for the Bulletin of the American Meteorological Society (BAMS). For GOZCARDS O_3 v2.20, the stratospheric retrieval pressure grid is twice as fine as for v1.01; there are now 12 regularly-spaced levels per decade change in log of pressure. ~~The~~ UARS MLS O_3 data were not included in v2.20, since these retrievals are not readily available on the finer vertical grid (although approximations such as interpolation could be used); also, there is no easy provision of UARS MLS retrieval uncertainties on a finer grid. The most significant change for the new merged O_3~~ozone~~ is the effect of using the ~~updated and~~ more robust version 7 data from SAGE II (Damadeo et al., 2013). Version 7 uses National Aeronautics and Space Administration (NASA) Global Modeling and Assimilation Office (GMAO) Modern-Era Retrospective Analysis for Research and Applications (MERRA) temperature (T) profile data (Rienecker et al., 2011) in the retrievals, and these values (rather than T from the National Centers for Environmental Prediction (NCEP)) also have a significant impact on the conversion of SAGE II O_3 from its native density/altitude grid to the GOZCARDS mixing ratio/pressure grid. Also, Aura MLS v4.2 O_3 data are now used (instead of MLS v2.2); HALOE v19 O_3 profiles are included, after interpolation to the finer pressure grid before merging. As a result of these change~~improvement~~s ~~in GOZCARDS O_3~~, we have observed closer agreement and larger correlation coefficients between the Stratospheric Water and Ozone Satellite Homogenized (SWOOSH) ozone data ~~record~~ (Davis et al., 2016) and GOZCARDS v2.20 O_3 ~~time~~ series than between SWOOSH and GOZCARDS v1.01 (SWOOSH ~~O_3~~ also uses SAGE II v7 O_3~~ozone~~ data).~~;~~ M~~m~~ore details regarding the impact of GOZCARDS v2.20 O_3 on trends are provided in

Sect. 5~~ and in the Appendix~~. We note that no ACE-FTS data were included in this newer version of GOZCARDS O_3.

3 WACCM (CESM1) description and simulations

WACCM (CESM1) is a chemistry climate model of the Earth's atmosphere, from the surface to the lower thermosphere (Garcia et al., 2007; Kinnison et al., 2007; Marsh et al., 2013; Garcia et al., 2017). WACCM is a superset of the Community Atmosphere Model, version 4 (CAM4), and includes all of the physical parameterizations of CAM4 (Neale et al., 2013) and a finite volume dynamical core (Lin, 2004) for the tracer advection. The horizontal resolution is 1.9° latitude x 2.5° longitude. The vertical resolution in the lower stratosphere ranges from 1.2 km near the tropopause to ~2 km near the stratopause; in the mesosphere and thermosphere the vertical resolution is ~3km. Simulations used here are based on the guidelines from the International Global Atmospheric Chemistry / Stratosphere-troposphere Processes And their role in Climate (IGAC/SPARC) Chemistry Climate Model Initiative (CCMI) (Morgenstern et al., 2017). Improvements in CESM1 (WACCM) for CCMI include a modification to the orographic gravity wave forcing, which reduced the cold bias in Antarctic polar temperatures (Garcia et al., 2017; Calvo et al., 2017) and updates to the stratospheric heterogeneous chemistry, which improved the representation of polar ozone depletion (Wegner et al., 2013; Solomon et al., 2015). In this work, there are two CCMI scenarios, spanning the 1990-2014 period. The first scenario follows the CCMI REF-C1 definition and three ensemble members were completed; this falls under the "free-running" scenario. We note that all the analyses herein are based on an average of these three simulations. We have checked that the three representations' departures from the average are small enough not to require separate comparisons for each case, when pursuing average or root mean square (RMS) differences versus observations, in comparison to differences using the 2nd model scenario (see below); this is also true for the model/data comparisons of RMS variability. This first model scenario includes forcing from greenhouse gases (CH_4, N_2O, and CO_2), organic halogens, volcanic aerosol surface area density and heating, and 11-year solar cycle variability. The sea surface temperatures are based on observations and the quasi-biennial oscillation (QBO) is nudged to observed monthly mean tropical winds over 86-4 hPa, as described in Matthes et al. (2010).

The second scenario is based on the CCMI REF-C1SD scenario and includes all the forcings of REF-C1, except for additional external QBO nudging. This scenario uses the specified dynamics (SD) option in WACCM (Lamarque et al., 2012). Here, temperature, zonal and meridional winds, and surface pressure are used to drive the physical parameterization controlling boundary layer exchanges, advective and convective transport, and the hydrological cycle. The meteorological analyses are taken from MERRA and the nudging approach is described in Kunz et al. (2011). The QBO circulation is inherent in the MERRA meteorological fields and is therefore synchronized with that in the "real" atmosphere. The horizontal resolution is the same as the REF-C1 version and the vertical resolution follows the MERRA reanalysis (from ~1 km resolution near the tropopause to about 2 km near the stratopause). The model meteorological fields are nudged from the surface to 50 km; above 60 km, these fields are fully interactive, with a linear transition in between.

Both WACCM versions used here contain an identical representation of tropospheric and stratospheric chemistry (Kinnison et al., 2007; Tilmes et al., 2016). The species included in this mechanism are contained within the Ox, NOx, HOx, ClOx, and BrOx chemical families, along with CH_4 and its degradation products. In addition, 20 primary non-methane hydrocarbons and related oxygenated organic compounds are represented, along with their surface emissions. In total there are 183 species and 472 chemical reactions; this includes 17 heterogeneous reactions on multiple aerosol types (i.e., sulfate, nitric acid trihydrate, and water-ice). For this work, the CESM1 (WACCM) REF-C1 and REF-C1SD simulations will generally be referred to as FR-WACCM and SD-WACCM, respectively. While the runs were originally designed to stop at the end of 2010, for this work, the forcing inputs have been extended through 2014.

4 Climatological comparisons and biases

We first describe some of the major climatological features for the various stratospheric species mentioned in the Introduction. We focus here on the main differences between average model trace gas abundances fromin the FR-WACCM and SD-WACCM runs and the corresponding dataabundances from Aura MLS for 2005 through 2014; this includes a sub-section on annual and semi-annual variations. Further analyses of interannual variations and trends are discussed in Sect. 5.

4.1 Average abundances

We provide climatological latitude/pressure contour plots in the Supplement (Fig. S1 for O_3 and H_2O, and Fig. S2 for HCl, HNO_3, and N_2O), showing Aura MLS and WACCM average mixing ratio distributions, averaged forover 2005- through 2014. Since such plots do not easily allow one to quantify the largest areas of model/data disagreement, we show in Fig. 1 (left column, top two panels) the percent differences between WACCM and the MLS climatologies; a positive value means that, on average, for 2005-2014,, the model values exceed the data values. Also, we show in the right column (top two panels) the absolute value of the average model/data difference divided by systematic error estimates (2σ values) for MLS O_3ozone. These error estimates have been provided in past MLS validation and error characterization work as tabulated "typical" (global average) profiles versusas a function of pressure; the latest update of such error estimates for version 4 MLS data areis provided by the MLS team in Livesey et al. (2018). The vertical profile of estimated systematic errors for MLS O_3ozone systematic errors is given in the Supplement (Fig. S3). Note that the MLS team can provide further systematic error details for MLS data users (e.g., for pressure levels not listed in the standard Tables). Past validation references for MLS O_3ozone include Jiang et al. (2007), Froidevaux et al. (2008a), and Livesey et al. (2008), as well as the more recent work covering many satellite (and other) instruments by Hubert et al. (2016). The original MLS data validation work for H_2O is from Read et al. (2007) and Lambert et al. (2007), who also described N_2O validation; MLS , whereas HNO_3 validation work was provideddiscussed by Santee et al. (2007).

The two bottom panels of Fig. 1 show a comparison between the two model runs (with percent differences on the left and the ratio of the absolute differences versus data on the right). The main conclusion from Fig. 1 is that Mmost of the model O_3 ozone climatology falls within about 5 to 10% of the data climatology, except in the upper troposphere and lower stratosphere (UTLS), where the SD-WACCM O_3 values are even larger than those from FR-WACCM (see bottom left panel). Our work focuses on the stratosphere, but both FR-WACCM and SD-WACCM average O_3 values at low latitudes are lower than the observed mean values from 215 to 261 hPa (and theseMLS retrieval pressure levels that lie in the upper troposphere at these latitudes). While the right column of Fig. 1 indicates that the difference to error ratio does not show a very significant systematic difference in this region, there are known MLS positive biases versus tropical ozonesonde data; , and this could account for at least part of the apparent model low bias, at least for the 215 hPa level; between this level and the tropopause (near 100 hPa), the SD-WACCM model values appear to be biased somewhat high. HoweverOn the other hand, O_3 mid-latitude O_3values from 100 to 215 hPa isare biased high in SD-WACCM in particular, with a difference to systematic error ratio larger than 2- to 3 in most of this region; otherwise, Fig. 1 shows that these ratios are usutypically less than 1 to 1.5. An illustration of the more significant differences is given in Fig. 2, for the O_3 data and model time series at 215 hPa for 50°N-60°N (which corresponds to these abundances represent lower stratospheric values). This shows that both FR-WACCM and SD-WACCM average values are larger than the data therein this region, and more so for SD-WACCM, for which the overestimate can be larger than 50%. In relation to this model overestimate of ozone, Imai et al. (2013) have showedn that the SD-WACCM O_3 values are also larger than those

~~from the~~ Superconducting Submillimeter-Wave Limb-Emission Sounder (SMILES) O_3 in the lowest portion of the stratosphere (~~near~~ 18-20 km). In Fig. 2, we also see~~observe~~ that the model O_3~~ozone~~ annual amplitude (AO) is larger than the observed amplitude; both models overestimate seasonal amplitudes by ~60%. W~~w~~e discuss the ~~stratospheric~~ AO more generally in the next section. If anything, MLS O_3 at mid-latitudes is ~~biased~~ slightly high (by ~ ~~roughly~~ 5%) with respect to a multi-instrument mean ~~ozone field~~ based on a large number of ~~monthly zonal mean~~ satellite data sets from the Stratosphere-Troposphere Processes And their Role in Climate (SPARC) Data Initiative (DI), as discussed by Tegtmeier et al. (2013), who also showed that the MLS O_3 seasonal cycle at mid-latitudes and 200 hPa is in ~~very~~ good agreement with the multi-instrument mean. The ~~limb emission measurements from~~ Aura MLS measurements for the species discussed here provide generally strongly peaked vertical averaging kernels with a resolution of 2.5-4 km (see Livesey et al., 2018, for sample kernel plots ~~of these kernels~~). **We have confirmed that smoothing the model profiles using the MLS averaging kernels (and a priori information) gives very little change (less than a few % ~~%~~ for the Fig. 2 example, and even less at higher altitudes) in O_3 abundances and seasonal cycles, and we see no real ~~real~~ need to use such ~~a~~ smoothing for our~~the model/data~~ comparisons~~ herein~~.~~;~~ T~~t~~he significant model/data differences in Figs. 1 and 2 are ~~thus~~ not caused by this sort of issue**. The mid-latitude O_3 differences mentioned above require more detailed investigations, but appear to implicate (in part) a transport-related issue in these models.

Figure 3 is the H_2O analog of Fig. 1, but for pressures reaching up to 0.01 hPa~~, since the Aura MLS H_2O retrievals cover both the stratosphere and the mesosphere~~; again, a typical profile of MLS systematic error estimates (Livesey et al., 2018) w~~h~~as ~~been~~ used in deriving the results ~~shown~~ in the top two right panels~~, which test for the existence of a significant bias between model and observations~~. We note that MLS H_2O v~~ersion~~ 3 stratospheric data ~~generally~~ exhibit a slight high bias (of a few to 5%) versus multi-instrument mean ~~value~~s, with a somewhat larger positive bias (of ~10%) in the lower mesosphere (Hegglin et al., 2013). Such biases are within the expected measurement systematic errors.~~;~~ MLS version 4 stratospheric H_2O data show essentially no systematic change versus~~in comparison to~~ version 3 (Livesey et al., 2018). FR-WACCM and SD-WACCM H_2O mean values are on the low side (by ~5-15%) relative to MLS H_2O in the upper stratosphere and in most of the mesosphere (see Fig. 3), implying that the models are in good agreement with the SPARC DI multi-instrument mean H_2O. There is independent evidence that MLS H_2O has a dry bias near the hygropause (at the low end of the vertical range shown in Fig. 3, where the bottom level is 150 hPa); this has been known for some time (~~see~~ Read et al., 2007; Vömel et al., 2007). This is also consistent with the existence of a model high bias relative to MLS near 150 hPa~~, assuming that the models come close to representing the H_2O climatology in this region~~. In terms of the significance of the biases, the right top two panels of Fig. 3 indicate that the H_2O model/data comparisons are generally in agreement within the estimated (2σ) systematic errors, and that this level of agreement is slightly better in the case of SD-WACCM.

For HCl, the climatological comparisons of Fig. 4 show that both models exhibit a small (5-10%) low bias versus MLS HCl in much of the stratosphere, with a stronger negative model bias in the tropic~~sal region~~ between 100 and 150 hPa. The model/data relative biases in stratospheric HCl are generally within the MLS HCl systematic errors (see Fig. 4, top two panels at right). MLS HCl is slightly on the high side of ~~the~~ multi-instrument mean climatological results provided in the SPARC DI report (SPARC, 2017). The small negative model bias in the upper stratosphere could also arise from the lack of a sufficiently pronounced decrease in upper stratospheric MLS HCl, as a result of the interruption in the main MLS HCl ~~target band~~ (band 13) data after early 2006 (~~see~~ Livesey et al., 2018). There is also a known ~~strong~~ positive ~~systematic~~ bias in MLS tropical HCl at 150 hPa ~~in the tropics~~ (~~see~~ Froidevaux et al., 2008b), so model underestimates in this region are not a sign of model weakness. We also note that both models exhibit a systematic difference versus HCl observations in the lower stratosphere (with larger differences for SD-WACCM), as well as a downward sloping pattern (equator to pole) in the southern hemisphere (SH), and smaller mean differences (for SD-WACCM) in the northern hemisphere (NH).

Figure 5 provides average comparisons for N_2O; this species is long-lived in the lower stratosphere, which means that good (or poor) model/data agreement in this region can confirm (or deny) accurate model representations of the dynamics. ~~(a good dynamical tracer).~~ While the mean lower stratospheric SH N_2O values are larger for SD-WACCM than for FR-WACCM (bottom left panel of Fig. 5), the mean absolute fit for SD-WACCM versus MLS is not significantly better. The most significant climatological differences with respect to the error bars (top two right panels in Fig. 5) are in the upper stratosphere at low latitudes; in this region, SD-WACCM agrees somewhat better with MLS. However, this is also where the mean abundances decline rapidly with height towards the limits of the MLS sensitivity. Not too surprisingly, this is also where the SPARC DI results for N_2O show the largest scatter in terms of percent differences (often exceeding 10-20%, see SPARC, 2017).

Finally, the ~~climatological~~ comparisons for HNO_3 (Fig. 6) reveal very few areas of ~~mean~~ model/data disagreements ~~significantly~~ outside the systematic uncertainties. However, there is a ~~general~~ model underestimation by both WACCM versions, especially in the polar upper stratosphere; this will be discussed more later. ~~We will see later that there are large recurring model/data differences in the upper stratosphere during certain months.~~ At high latitudes in the lower ~~portion of the~~ stratosphere, the models tend to overestimate the data. The upper troposphere is where the satellite-based HNO_3 data have been validated the least, but there is some evidence for a high MLS bias in this region~~, based on SPARC DI results~~ (see SPARC, 2017). While this might explain, ~~at least~~ qualitatively, why the models ~~are observed to~~ underestimate MLS tropical UT HNO_3 (see Fig. 6), more work is needed to better evaluate HNO_3 from models and satellite-derived data~~sets~~ in this particular region.

It is also worth emphasizing differences between ~~interesting to point out~~ modeled and observed seasonal changes in the polar lower stratosphere over Antarctica. ~~Figure 7 depicts average seasonal changes over the 70°S-80°S region at 46 hPa for 2005-2014, using Aura MLS observations and model values for comparison. We wish to emphasize the slope of the early winter decline in HCl, and~~ Indeed, we see in Figure 7 that ~~the~~ model HCl values at 46 hPa for 70°S-80°S do not decline as fast in early winter as shown~~indicated by~~ in the data, even though SD-WACCM tracks the interannual variability better than FR-WACCM does (see Fig. S4 for the relevant time series ~~from 2005 through 2014~~). ~~Many of the options and~~ Uncertainties regarding lower stratospheric heterogeneous chemistry modeling for SD-WACCM at high latitudes in the polar winter/spring have been discussed by Solomon et al. (2015)~~, who have pointed,~~ for one specific year (2011), including most of the features shown~~depicted~~ in Fig. 7. For our broader ~~time~~ period, we see that the average~~HCl~~ rate of HCl decline from May to July (dominated by nighttime conditions) is slower in both ~~the~~ SD-WACCM and FR-WACCM ~~results~~ than the corresponding mean HCl rate of change from MLS (top left panel of Fig. 7). Grooß et al. (2018) ~~have~~ recently discussed this HCl model/data discrepancy for dark polar vortex conditions, ~~including the potential impact of numerical diffusion issues in the Eulerian models,~~ in comparison to ~~their~~ simulations using the Chemical Lagrangian Model of the Stratosphere (CLAMS), which shows even larger HCl discrepancies. These authors discuss ~~some~~ possible mechanisms and uncertainties, and they argue that additional decomposition of condensed-phase HNO_3 might play a role, possibly via galactic cosmic ray impacts. Since this rapid decline in HCl occurs during polar night, ~~these authors point out that~~ this ~~early~~ HCl issue does not lead to much difference in polar ozone loss rates, which only become significant during sunlit conditions (early spring). Figure 7 confirms that, on average, the SD-WACCM O_3 decline and rise match the data well. We also note that FR-WACCM shows smaller-than-observed declines in HNO_3 and H_2O, whereas SD-WACCM matches these observations much better. The temperature panel (bottom center) gives a possible reason for these differences, as T from FR-WACCM is larger by a few degrees during the coldest phase than T from SD-WACCM (MERRA-based), and larger than the MLS-derived values. This would lead to less irreversible denitrification and dehydration. The ~~general~~ nature of the Fig. 7 results ~~shown in Fig. 7~~ is similar at other lower stratospheric pressures ~~(and for latitudes poleward of 80°S)~~, although there is some variability in the magnitude of the differences. Over the Arctic region (not shown here), temperature-related differences are not as ~~systematic or as~~ large as over Antarctica, but similar model/data differences in the early winter rate of HCl decline ~~in HCl~~ exist there also (as mentioned by Grooß et al., 2018).

As an addendum regarding the evaluation of models in comparison to data sets, we provide in Appendix A1 the results of a model grading approach ~~that has been~~ used before~~in the past~~ (e.g., Douglass et al., 1999, Waugh and Eyring, 2008). We find ~~(see Appendix A1)~~ that this grading method often leads to low grades (see Fig. A1), if applied using systematic uncertainty estimates from the MLS ~~data characterization work~~ team (Livesey et al., 2018). The model results, as good as they are in many respects, cannot always match the data closely enough, at least based on such grades (although th~~e~~ise grading formulation could ~~also~~ also be reconsidered). As mentioned in the Appendix, the multiplicative error factor (see Equation A1) can be increased (e.g., from 2 to 4) to force these grades (see Fig. A2) to span a more useful range (in the plots). Indeed, we observe similarities between Fig. A2 (top two panels) and Fig. 1 (middle panels): poorer SD-WACCM performance for pressures ≥ 100 hPa, better mid-latitude results near 30-40 hPa, and poorer performance again near 3 to 5 hPa. Also, we observe the poorest H_2O grades near 1 hPa in Fig. A2 (bottom panels), which is similar to the poorer performance in the middle panel of Fig. 3, while the best stratospheric H_2O grades are found near 10-20 hPa, which matches the best performance (smallest values) in Fig. 3 (top two panels at right). In the case of merged data records, ~~we note that~~ it is generally more difficult to estimate systematic errors. The GOZCARDS data analyses led ~~to conservative (i.e. to~~ possibly somewhat pessimistic~~)~~ systematic error estimates ~~as a function of latitude and pressure~~ (see Froidevaux et al., 2015), and these ~~that~~ are significantly larger than ~~the systematic~~ error estimates for MLS data only. ~~There are~~ O~~o~~ther methods ~~that~~ could lead to useful error estimates, through the use of multi-satellite data sets and the spread between these (see SPARC, 2017), for some species at least. Moreover, when one considers ~~relative variations such as~~ anomaly time series (as done in a subsequent section), it becomes even less clear how to best assign uncertainties in the context of "error-weighted" grades.~~;~~ S~~s~~ome data records may also drift with respect to others, or with respect to ground-based data, so that the actual errors will change with time (and possibly ~~with~~ location ~~as well~~)~~, in a difficult to determine way~~. We do not pursue this more traditional grading approach further here, especially as the two models we are comparing often lie fairly close together. For multi-model comparisons (which is outside the scope of this work), one could consider how to best apply grading methods such as the one in Appendix A1, along with other diagnostics such as those discussed here; in the end, the most important aspect of such analyses may~~possibly~~ lie~~s~~ in the relative values of ~~the~~ grades or diagnostics for different models.

4.2 Annual and semi-annual cycles

Figure 8 displays the amplitudes of annual and semi-annual variations for MLS O_3 ~~ozone~~ and ~~the corresponding FR-~~WACCM ~~and SD-WACCM~~ runs for 2005-2014. ~~We obtained~~ T~~t~~hese results come from a simple regression fit to the monthly mean ~~time~~ series in each latitude/pressure bin. The primary time dependence of the fitted function is given by additive sine and cosine terms (with 12-month and 6-month periods), in addition to constant and linear trend terms; the AO and SAO amplitudes are given by the square root of the sum of the squares of the corresponding fitted coefficients. We see from Fig. 8 that the overall data and model patterns of AO and SAO variability are quite similar. The O_3~~ozone~~ AO amplitudes peak ~~(in ppmv)~~ at mid- to upper stratospheric levels, with high latitude variations also observed as a result of the effects of winter/spring polar chemistry and dynamics.~~;~~ T~~t~~he lower stratospheric peak AO amplitudes are more prominent over the southern polar regions, where stronger O_3 depletion occurs on a seasonal basis. These MLS O_3 AO patterns are ~~very~~ similar to those obtained by Schoeberl et al. (2008), using a much shorter ~~time~~ period (Sep. 2004 to Dec. 2006); the same holds for other species (H_2O and HCl) ~~considered in that work and here~~. The observed SAO amplitude for O_3 exhibits strong peaks in the upper stratosphere, both in the tropics and at high latitudes. The anti-correlation between O_3 and temperature as a result of temperature-dependent photochemical production and loss terms for O_3 has long been known to cause most of the O_3 variability in the upper stratosphere (~~see~~ Perliski et al., 1989~~,~~

~~in relation to the AO and the SAO~~). The AO and SAO amplitudes obtained in that and other past studies (e.g., ~~see~~ Ray et al., 1994) are ~~very~~ similar to the ~~amplitude~~ patterns shown here. If we look more closely (and based on AO amplitude ratio plots not shown here), there are often O_3 AO amplitudes 20-80% larger than those derived from MLS data for both WACCM runs in the lower stratosphere (from 50 hPa at low latitudes to 215 hPa at high latitudes).~~;~~ S~~s~~uch~~ a~~ model overestimates of the AO amplitude were~~as~~ shown in ~~the time series example of~~ Fig. 2. ~~More generally, O~~Outside of this region, we observe somewhat closer fits to the MLS AO amplitudes for the SD-WACCM version, and~~but~~ both models track the data and each other well, with AO amplitudes typically within a range of ~25%~~ of the MLS AO amplitudes~~. For the Antarctic lower stratosphere, the timing and magnitude of the seasonal recovery after the ozone hole plays a role, and ~~in this respect,~~ we have observed that SD-WACCM generally fits the MLS data better than FR-WACCM does. In the tropical upper stratosphere, where the SAO is larger than the AO (see Fig. 8), the SD-WACCM results match the observed SAO amplitudes slightly better than those from FR-WACCM. Despite the existence of a few model/data differences, these AO and SAO amplitude comparisons, coupled with our examination of model/data amplitude ratio plots as well as the time series (which include the phase information), do not elicit major concerns regarding the model characterization of the primary processes expected to govern these modes of O_3 variability. The study of dynamical forcing mechanisms in relation to such modes continues to be an active area of research (e.g., see Ern et al., 2015, and Smith et al., 2017).~~for a discussion of wave-driving and the SAO). Also, Smith et al. (2017) have recently shown that analyses of Aura MLS geopotential height data lead to derived tropical zonal mean winds that agree well with those derived from Sounding of the Atmosphere using Broadband Emission Radiometry (SABER) geopotential heights, and with direct wind data, thus enhancing our knowledge of tropical atmospheric dynamics (including the SAO and QBO).~~

For H_2O, a similar overview of the AO and SAO amplitudes is given in Fig. 9, which covers the ~~vertical~~ range from 100 to 0.01 hPa. A peak in these amplitudes resulting from ~~the~~ seasonal downward transport before the winter, followed by wintertime dehydration, is observed in the lower stratospheric southern polar region; we note that ~~the~~ SD-WACCM results match this feature better than~~ the~~ FR-WACCM ~~simulation~~ does. Other ~~interesting~~ features include the southern hemisphere's upper stratospheric AO peak in the extra-tropics~~al region~~. This has been seen by many satellite~~-based~~ measurements (see ~~, as discussed in the comparisons by~~ Lossow et al.,~~ (~~2017a). Lossow et al. (2017b) ~~have~~ explained this "island" feature in more detail~~, with the help of model simulations and analyses~~; they argue that vertical advection tied to the upper branch of the Brewer-Dobson circulation largely explains the seasonal highs (lows), via downwelling (upwelling). They also show that an AO maximum is observed as well in other species in roughly the same region, including in N_2O MIPAS data.~~;~~ W~~w~~e confirm this behavior (see also Fig. S5) from the N_2O AO amplitude feature observed in MLS data, as well as in the model~~WACCM~~ runs (and more so in SD-WACCM). The derived AO and SAO amplitude patterns in H_2O from Lossow et al. (2017a) are consistent with what we find~~show here~~; this includes the peak values in the upper stratosphere and mesosphere, attributed to ~~the~~ combined effects of photochemistry and vertical transport. For ozone, a dominant feature in the SAO amplitude exists in the tropical upper stratosphere; see Lossow et al. (2017a) for a brief review of past work explaining such dynamically-driven features for H_2O. While there is generally a good level of model/data agreement in the main H_2O AO and SAO patterns, both WACCM comparisons tend to underestimate observed AO and SAO amplitudes in the lower stratosphere and overestimate AO amplitudes in the SH mesosphere, while slightly underestimating ~~the~~ mesospheric SAO amplitudes in both polar regions. The largest amplitude differences reach a factor of two, in places, for the lower stratospheric model underestimates~~, which cannot be caused by slight model underestimates of average MLS H_2O (as seen in Fig. 3)~~; however, we note that this ~~the lower stratosphere~~ is also the region where AO and SAO amplitudes are smallest (typically < 0.1 ppmv).

~~Turning briefly to the other species discussed here, F~~For N_2O, we ~~have~~ already mentioned the existence of the upper stratospheric AO peak (the "island" feature described by Lossow et al., 2017b) in the southern hemispheric extra-tropical region.

This is similar to the H_2O AO amplitude maximum feature, but the N_2O seasonal variations are anti-correlated with H_2O, as demonstrated by Lossow et al. (2017b), using Michelson Interferometer for Passive Atmospheric Sounding ~~(~~MIPAS~~)~~ data (and as is also apparent in ~~the~~ MLS time series, not shown here). Furthermore, we observe in Fig. S5 a somewhat better match in the AO and SAO N_2O amplitude patterns for SD-WACCM than for FR-WACCM versus MLS, in particular for tropical to southern mid-latitudes. ~~We note that~~ T~~t~~his is also manifested in better time series fits for SD-WACCM, besides the closer match in average values; ~~as mentioned in the Introduction,~~ this is what one would generally expect f~~or~~om these two model versions. ~~We saw also in Fig. 5 that FR-WACCM overestimates the average values of upper stratospheric tropical N_2O.~~

For HCl and HNO_3, the AO and SAO amplitudes are dominated by the variations at high latitudes (see Figs. S6 and S7). The model HCl upper stratospheric AO and SAO amplitudes match up fairly well with the observed amplitudes, despite the aforementioned issues relating to MLS upper stratospheric HCl trends. ~~The NH lower stratospheric polar variations are somewhat underestimated by the models, and more so by FR-WACCM run, which (based on time series plots) exhibits larger values at the low end and smaller values at the high end of the HCl summer to winter cycle.~~ For HNO_3, the main AO and SAO model features follow the MLS patterns, although there is a model underestimation of the amplitudes in the upper stratospheric polar regions, because these models do not properly capture the observed recurrences of enhanced HNO_3, as mentioned earlier (see also Sec. 4).

5 Time series comparisons

5.1 Anomaly time series: fits and variability

5.1.1 Fits

We wish to evaluate which of the two WACCM models provides a better match, or fit, to the temporal variations in observed deseasonalized anomalies. We would expect SD-WACCM to generally fit these anomalies better than FR-WACCM. ~~We analyze the model and data deseasonalized anomaly time series, obtained by differencing each month's zonal mean value from the long-term averages for the same month.~~ We~~, then~~ calculate a diagnostic of model fit to the data, by using the RMS differences between these deseasonalized model and data series, and normalize by dividing this quantity by the RMS of the data anomalies themselves. A~~Thus, a~~ diagnostic value ~~that is~~ much less than unity means that the match to the ~~time~~ series is much smaller than the typical variability; this also implies a good fit to ~~the~~ observed anomalies. In ~~the~~ Appendix ~~(~~A2~~)~~, we provide the mathematical expression for this "RMS difference diagnostic". A better model fit to the observ~~ationsed~~ ~~anomaly series~~ will be given~~represented~~ by a smaller "RMS difference diagnostic" value. We also calculate the ~~standard~~ (Pearson) correlation coefficients, R, between model and data anomalies, and we use R^2 as another measure of "goodness of fit" for the models. The first diagnostic is unitless and does not depart too much from the 0 to 1 range; R^2 is limited to the 0 to 1 range, with larger values indicating a higher degree of linear correlation. We also use these two diagnostics together, by calculating the ratio of R^2 over the RMS difference diagnostic to obtain a "combined diagnostic". This diagnostic could have a large value (~~and~~ a good model result) from both a large R^2 ~~value (~~in the numerator~~)~~, meaning a high correlation with observed anomalies, and a small RMS difference ~~(~~in the denominator~~)~~, implying a good fit ~~to these observations~~. This ~~also~~ tends to amplify ~~the~~ differences between two model comparisons to the same data ~~series~~. An ideal model fit would correlate tightly in time to observed oscillations, but also exhibit the right magnitude ~~for these variations~~ by "hugging" the anomaly series ~~without being off in magnitude~~. Indeed, two model series could have oscillations in phase with data variations and ~~thus~~ the same R^2 values, but with different amplitudes ~~(~~and ~~different~~ overall fits~~)~~; conversely, two model series could have different R^2 values versus observations, if one is more out of phase than the other, but they could still produce similar RMS difference fits.

In Figure. 10 shows, we display latitude/pressure contours plots of the above diagnostics for FR-WACCM and SD-WACCM O_3 anomalies in relation to MLS anomalies for 2005-2014. We immediately see from the top two panels that the SD-WACCM RMS difference diagnostic values (in the 0.2 to 1 range in the stratosphere) are smaller than those from FR-WACCM (betweenwith typical values between about 0.8 and 1.2). Values of R^2 values, between about 0.7 and 0.95 in most of the stratosphere, (middle panels) show that SD-WACCM also correlates very well with the observations from 2005-2014;, R^2 with values typically between 0.7 and 0.95 in most of the stratosphere, and sis somewhat smallerpoorer correlations in the UTLS. The FR-WACCM ozone O_3 series tend to correlate fairly well with the data at low to mid-latitudes for pressure levels between about 70 and 7 hPa, which is also where the FR-WACCM RMS difference diagnostic shows better performance than in other regions, and (as an explanation) where the dynamics are nudged (to tropical winds) in a similar way as for SD-WACCM. However, FR-WACCM shows poor performance (almost zero correlation) at high latitudes and in all of the uppermost stratosphere.

Figure 11 shows sample O_3 series for the upper stratosphere (at 2.2 hPa) for 0°-10°N as well as 40°N-50°N. Although the tropical series show good correlations overall for both models versus data, some of the details in theof observed semi-annual peaks and the interplay between the AO, SAO, and the quasi-biennial oscillation (QBO) are better followed by the SD-WACCM curve. The differences in O_3 amplitude and phase are more clearly displayed in the bottom (left) panel, which showings the deseasonalized anomalies. Diagnostic values provided in this panel for both models distinctely show that SD-WACCM performs better than FR-WACCM here, with a much larger R^2 value and a smaller RMS difference diagnostic value, and hence, a much better (larger) combined diagnostic value. The same comments apply to the right two panels of Fig. 11, which showcase the NH mid-latitudes at 2.2 hPa. In the high latitude lower stratosphere, the poorer FR-WACCM results in Fig. 10 are generally caused by time series that are observed to be less in-phase with the polar winter/spring variations, as well as by more departures in the variation magnitudes of such variations; we will return later to sample polar time series for ozone and other species. The bottom two panels in Fig. 10 amplify the model differences between the models, with values of the combined diagnostic values mostly below 1 in most regions for FR-WACCM, but values between 1 and 3 for the whole stratosphere in the SD-WACCM case. The more realistic dynamics in SD-WACCM , coupled with the same chemistry as FR-WACCM, allow for better SD-WACCM fits to the data, as shown quantitatively in Figs. 10 and 11. These plots also point to poor upper tropospheric results for both models (with SD-WACCM slightly better)in the upper troposphere, albeit with somewhat better results for SD-WACCM, although this is not the focus of this paper.

We now turn to Figure. 12 for a describes iption of the samee diagnostics as above, but for model/data H_2O comparisons, and a top pressure level at 0.01 hPa. We observe, Aagain, that the diagnostics of fit are usually much better for SD-WACCM, which yields R^2 values of 0.6 to 0.9 and RMS difference diagnostic values below 1 for most of the stratosphere and lower mesosphere, and therefore better combined diagnostic results than FR-WACCM as well. The SD-WACCM diagnostics themselves are poorestworsen in the upper mesosphere and atin the high latitude regions near 215 hPa. In the upper stratosphere and upper mesosphere, the better diagnostic results for SD-WACCM are seen in the time series (not shown here) as a better match versus the MLS H_2O series anomalies in terms of the interannual variability at all latitudes, as well as for some seasonal peaks at high latitudes.; Wwe interpret this as the result of a better dynamical representation of the mesosphere for SD-WACCM. TWe note that the high-quality representation of mesospheric composition by SD-WACCM is also demonstrated in comparisons to measurements of CO profiles above Kiruna, Sweden, by MLS and the Kiruna Microwave Radiometer (Ryan et al., 2018). For H_2O near 200 hPa, poor fits at high latitudes occur where MLS data is known toexhibit have a significant dry bias versus sonde and Atmospheric Infrared Sounder (AIRS) data, as discussed previously in MLS data documentation (Livesey et al., 2018), as well as by Vömel et al. (2007) and Davis et al. (2016). MLS H_2O is low by a factor of several here versus the WACCM runs (which show values of 10-60 ppmv). The data variability may also be affected by the dry bias retrieval (and oscillation) issues at the lowest altitudes at~~for these~~ high latitudes~~ regions~~, where observed anomalies are more poorly tracked by the models; a planned future update to the MLS H_2O retrievals will~~might~~ help to mitigate this discrepancy.

~~Figure 13~~ Figure S8 displays results similar to Fig. 10 but for stratospheric HCl. SD-WACCM HCl results versus MLS are ~~also~~ ~~generally~~ superior to those from FR-WACCM and show high correlations ($R^2 > 0.7$) in most of the lower stratosphere, with somewhat poorer results in the upper stratosphere, where the RMS difference fits as well as the combined diagnostic are poor for both models. ~~We trace~~ ~~T~~this~~e~~ upper stratospheric issue is caused by~~ack to a~~ a ~~known~~ data problem in this region, where MLS HCl trends are known to be too flat (close to zero)~~ and the models depart more from the observations (down to about 10 hPa), notably in the RMS difference diagnostic~~. Poorer correlations are observed for FR-WACCM at high latitudes, ~~which we will return to later for the lower stratosphere,~~ but even~~ fairly~~ subtle differences in the timing (phase) can lead to~~ significantly~~ poorer correlations. P~~The p~~oorer fits at 100 hPa in the deep tropics are related to a large underestimate of the ~~mean~~ data, which may be caused,~~ at least~~ in part, by a~~n MLS~~ high bias from MLS (Froidevaux et al., 2008b), but more so, for R^2, by ~~some~~ out-of-phase model variability~~ as well~~. ~~It will be difficult to resolve this issue until more realistic (lower) MLS HCl values are obtained, and as the phasing may also change with new retrievals.~~

For ~~the dynamical tracer~~ N_2O, we also observe in Fig. S9~~14~~ (showing model comparisons to the 68 to 1 hPa observations from the 190 GHz MLS N_2O band for 2005-2014) that SD-WACCM fits the data better than FR-WACCM, in both the R^2 and the RMS difference categories. FR-WACCM exhibits poor results in the upper stratosphere and at high latitudes in the lower stratosphere. Both models exhibit poorer RMS fits and poorer correlations in the tropical lower stratosphere. Partly, this appears to be caused by a model underestimation of the MLS N_2O variability in this region, with some QBO phasing differences as well.

HNO_3 results (see Fig. S10)~~Fig. 15)~~ show, again, better fits to ~~the~~ stratospheric MLS data from SD-WACCM than from FR-WACCM, and poor performance from FR-WACCM at high latitudes. Both models do poorly in the upper stratosphere, and Fig. 13~~6~~ illustrates the magnitude of this discrepancy in the region (3.2 hPa and 70°S-80°S) where it reaches its maximum~~, in terms of mixing ratio values~~. Since its launch, MLS has been observing very large values of HNO_3 in the upper stratosphere, mostly in the polar regions during winter. The WACCM runs used here do not ~~have the right chemistry in the mesosphere and upper stratosphere~~ include the necessary photochemical pathways, including the effects of energetic particle precipitation on ion chemistry in the upper atmosphere, to adequately represent such variations; implementation of the ~~many~~ necessary missing chemical reactions has not made its way into most CCMs. The solution seems~~ is believed~~ ~~to be~~ tied to ion cluster chemistry during energetic particle precipitation (EPP) events, which includes large solar proton events (SPEs) as well as more regular auroral~~l-type~~ activity. ~~The study of upper stratospheric NOx enhancements tied to auroral activity and other EPP events has a long history based on other satellite measurements and modeling (e.g., Kawa et al., 1995; Callis and Lambeth, 1998; Siskind et al., 2000; Orsolini et al., 2005; Randall et al., 2007; Reddmann et al., 2010). Yearly upper stratospheric enhancements in ground-based microwave retrievals of HNO_3 profiles over Antarctica were discussed by de Zafra et al. (1997) and de Zafra and Smyshlaev (2001).~~ Direct high altitude EPP effects enhance NOx, which can propagate downward in polar winter and increase stratospheric NOx and HNO_3 via this indirect effect and conversion of N_2O_5 on ion water clusters (Böhringer et al., 1983). Large polar enhancements in upper stratospheric HNO_3 ~~(and other species)~~ were observed by the ~~Michelson Interferometer for Passive Atmospheric Sounding (~~MIPAS~~)~~ after ~~significant~~ SPE activity in 2003~~, as presented in several papers~~ (Orsolini et al., 2005; von Clarmann et al., 2005; Lopez-Puertas et al., 2005; Stiller et al., 2005). More complex modeling ~~using modified chemistry and transport-related effects~~ (e.g., ~~e.g., Jackman et al., 2008;~~ Funke et al., 2011; Verronen et al., 2011; Kvissel et al., 2012; Andersson et al., 2016) has produced EPP-induced enhancements in high latitude HNO_3, with related improvements in model/data comparisons into the mesosphere~~, and in comparisons of other species~~. Regarding ~~the~~ low latitude upper stratospheric HNO_3

comparisons, the poorer model fits (even for SD-WACCM) seem to be caused at least in part by more noisy and variable MLS data, under low HNO_3 conditions. Finally, tropical MLS HNO_3 data at 147 and 215 hPa are not fit well by either model, as the data exhibit significant seasonal oscillations between 0.2 and 0.5 ppbv (with larger amplitudes occurring at 147 hPa), whereas model values are smaller than 0.1 ppbv. There have been very few tropical UT validation comparisons for HNO_3 (see Santee et al., 2007), but in situ HNO_3 data from an airborne chemical ionization mass spectrometer hav showne indicated that UT HNO_3 tropical valuesmixing ratios are mostly below 0.1-0.2 ppbv (Popp et al., 2007, 2009).

We saw in Sect. 4 that there are generally good climatological comparisons between SD-WACCM and MLS variations over the Antarctica region during polar winter/spring, except for the rate of HCl decline in during early winter; also, poorer results are obtained by FR-WACCM. We also also find, (not too surprisingly,) that interannual differences in lower stratospheric chemical evolution over Antarctica are not as faithfully reproduced by FR-WACCM as by SD-WACCM, although we do not show more related details here. This is shown by the anomaly time series comparisons of For a more in-depth look at the lower stratosphere over Antarctica, Fig. S11Fig. 17 _displays anomaly time series comparisons for O_3 and temperature at 68 hPa and 70°S-80°S, along with the associated model diagnostic values. , along with the diagnostic quantities that we are using. We note (top panel) the poorer O_3 correlations for FR-WACCM ($R^2 = 0.22$) than for SD-WACCM ($R^2 = 0.89$), as well as the poorer RMS difference values (0.89 for FR-WACCM versus 0.33 for SD-WACCM), with correspondingly poorer results in the FR-WACCM combined diagnostic (0.25) versus SD-WACCM (2.67). Similar differences in the diagnostics are obtained in the bottom panel for temperature (T) anomalies, showing excellent agreement between SD-WACCM and retrieved temperature anomaly time series from MLS, both for the R^2 and RMS fit diagnostics. One should not expect the free-running model, even if it has certain tropical QBO-related constraints that mimic those from the SD version, to perform as well in terms of predicted high latitude temperatures as the SD-WACCM run, driven by realistic (MERRA) meteorological winds, as well as temperatures, which match up closely to the observed temperatures from MLS. These plots also show that springtime anomalies dominate the variability, with warmer than usual springs (induring October in particular), such as in 2012 and 2013, leading to more positive ozone anomalies, i.e. less ozone depletion; conversely, years (2006, 2008, 2010, 2011) with colder than usual springstime conditions are correlated with negative ozone anomalies and more depleted conditions. Other factors (besides local mean temperatures) can significantly influence interannual variability and longer-term ozone loss over Antarctica; this includes the strength of the vortex, total chlorine abundances, the phase of the QBO, tropospheric wave driving, the timing of warming events, and the impact of aerosols (e.g., Scaife et al, 2005; Parrondo et al., 2014; Strahan et al., 2015; Langematz et al., 2016; Solomon et al., 2016). We saw in Sect. 4 that there are generally good comparisons between SD-WACCM and MLS variations over the Antarctic region during polar winter/spring, except for the rate of HCl decline during early winter; also, poorer results are obtained by FR-WACCM. We also find (not too surprisingly) that interannual differences in lower stratospheric chemical evolution over Antarctica are not as faithfully reproduced by FR-WACCM as by SD-WACCM, although we do not show more related details here.

5.1.2 Variability

Given our expectations that SD-WACCM would match observed time series of multiple species up better than FR-WACCM to the observed time series of multiple species, and having demonstrated this in the previous section, we turn to what should be a more fair comparison between the two sets of model results, namely the variability aspect. We calculate the ratio of model to data interannual variability, as obtained from the root mean square values of deseasonalized monthly anomaly time series, expressed as a percent of climatological (full period) means; a simple linear trend is first subtracted from the time series, so that the variability comparisons remove any significant trend differences. We do this for the MLS data considered previously (starting in

~~2005)~~, but also for longer-term time series, using the GOZCARDS data~~ records~~. The models are sampled ~~monthly ~~following the monthly sampling of the data sets (but not at the daily sampling level of detail); sampling plays a role for the longer-term (merged) GOZCARDS data, which are comprised of some unevenly-sampled occultation data records (depending on latitude and pressure). Figure 14 ~~18 ~~compares the O_3~~ozone~~ variability ratios ~~(for ~~models versus data) using as a reference the MLS 2005-2014 data (top two panels), and the GOZCARDS merged ozone (1992-2003) data (bottom two panels). To first order, we observe similar patterns for both time periods. The SD-WACCM variability is generally within 10-20% of the data variability (ratio values between 0.8 and 1.2). The FR-WACCM variability is ~~generally ~~somewhat smaller than the data variability in the polar regions and in the upper stratosphere. ~~In a recent study, ~~Bandoro et al. (2017) also found that the free-running version of WACCM displays ~~somewhat ~~smaller ozone variability in the upper stratosphere, both for shorter-term and longer-term variabilities, than the observed variability, based on the merged SWOOSH O_3 data record. In the lower stratosphere at low to mid-latitudes, ~~we observe that ~~FR-WACCM exhibits slightly larger variability than the data, whereas SD-WACCM shows slightly smaller variability than the data; Bandoro et al. (2017) found that FR-WACCM slightly overestimates the decadal variability in this region. ~~If a free-running model exhibits significant differences versus observed variability, this has some implications regarding predictions of trend detection feasability, as free-running models are used for such predictions; indeed, trend error bars increase if the variability increases. ~~We also see in Fig. 14~~18~~ that at high latitudes, FR-WACCM ~~tends to~~ underestimates the actual variability, ~~which means that interannual swings in lower stratospheric ozone, in particular, are more muted in the model, ~~whereas this is less of an issue for SD-WACCM, with its more realistic representation of the dynamics; as an example, refer to Fig. 13~~6~~, for model and data anomalies at 68 hPa and 70°S-80°S.

A similar overview of model/data variability ratios is provided for H_2O in Fig. 15~~19~~, for~~which covers~~ ~~both ~~the mesosphere and the stratosphere. Here~~In this case~~, while the variability from SD-WACCM is ~~somewhat ~~closer than that from FR-WACCM to the data variability during both ~~time ~~periods, the tendency for both models is to underestimate the observed variability, with FR-WACCM show~~exhibit~~ing a stronger underestimate in the upper stratosphere and mesosphere. Such a~~A~~n~~ variability~~ underestimate for FR-WACCM implies that a~~ny~~ trend detection in the future will require more years of data~~ in the real atmosphere~~, if H_2O continues to have larger variability than model~~ expectation~~s. ~~Also, this will translate into smaller estimated uncertainties in the model-derived trends in comparison to the observations (as we will see in the next section on trends). ~~The FR-WACCM underestimate of the variability is sometimes by as much as a factor of two, although it is more typically by ~30% (see Fig. 15~~19~~). For a ~~time ~~series with RMS variability about the fit represented by σ_t, the number of years needed to statistically detect a trend is proportional to $\sigma_t^{2/3}$ (Weatherhead et al., 1998), and thus, an increase of σ_t by factors of 1.3, 1.5, and 2.0, for example, will lead to an increase in the number of years for trend detection by factors close to 1.2, 1.3, and 1.6, respectively. In the tropical lower stratospheric case, H_2O and temperature values and anomalies for 1992-2014 are shown for 100 hPa and 10ºS-20ºS in Fig. 16.~~20.~~ Again, we note the smaller-than-observed variability in ~~the ~~model H_2O oscillations, with SD-WACCM tracking the data better. This correlates with the temperature series, where smaller variability is seen in FR-WACCM, in comparison to SD-WACCM (which follows ~~the ~~MERRA temperatures); we also note that FR-WACCM temperatures are somewhat larger (by ~1K on average) than SD-WACCM temperatures in this region. This poorer tracking of cold point temperatures for FR-WACCM (Fig. 16) has likely implications for poorer stratospheric trend results for FR-WACCM H_2O as well, as we will actually observe in the trends section. It is well known that stratospheric entry level H_2O is governed by temperatures near the tropopause "cold trap"; the monthly average variations shown here are similar to what has been shown in past H_2O work (e.g., Randel et al., 2004, 2006; Randel and Jensen, 2013). Brinkop et al. (2016) used model runs from both free-running and nudged simulations to analyze the impacts of different constraints, including sea surface temperatures (SST) and meteorological fields, on "sudden" drops in H_2O~~water vapor~~; they found that several of these factors play a role in the H_2O

variations, including the timing of ENSO and SST variability, the phasing with the QBO, cold point temperatures, as well as the correct dynamical model state. Many other analyses of the relation between entry level H_2O, tropopause temperatures, transport, and convection have been carried out previously (e.g., Holton and Gettelman, 2001; Jensen and Pfister, 2004; Fueglistaler and Haynes, 2005; Rosenlof and Reid, 2008; Read et al., 2008; Schoeberl et al., 2013). Our point here is that the WACCM H_2O anomaly series underestimate the observed variability. We note also that this model underestimate exists if we calculate relative variability using a maximum minus minimum range from yearly average anomalies rather than monthly averages.~~; the resulting variability ratio patterns (not shown) are similar, overall, to the contour plots of Fig. 19.~~ We provide a global view of lower stratospheric variability differences (models versus data) in the anomaly ~~time~~ series comparisons at 83 hPa for all latitude bins in Fig. S12~~8~~. This also shows that the observed interannual changes in H_2O are ~~clearly~~ better followed by ~~the~~ SD-WACCM ~~time series~~ than by FR-WACCM, including the drop in H_2O after 2011 (see Urban et al., 2014). While lower stratospheric H_2O variability is underestimated by SD-WACCM by ~20%, the ~~actual~~ correlation between SD-WACCM and ~~the~~ observed anomalies is very good (as was shown in Fig. 12).

For HCl, we show the (detrended) variability ratios in Fig. 17~~21~~. The observed HCl variability is fairly well matched (within ~20%) by both models in the MLS time period, with an edge given to SD-WACCM. The observed variability is often underestimated (by ~30%) by both FR-WACCM and SD-WACCM in the earlier period (1992-2003). We believe that the HALOE sampling plays a role in this, i.e. even if we limit the model comparison (as we do here) to just the same months as when HALOE observations occurred, incomplete sampling in latitude and time can lead to differences versus a fully sampled model (see Toohey et al., 2013), and more so in the polar regions where the HCl variability is large. In the upper stratosphere, the variability ratios are comparable to or somewhat smaller than those in the middle stratosphere, and there is a 20-30% underestimate of the observed variability, which is based on HALOE HCl observations for ~~the~~ 1992-2003 ~~period~~. There have been difficulties in fully understanding (or modeling) observed upper stratospheric HCl variations before the declining phase that started after about 2000 (Waugh et al., 2001); see also Sect. 5. For 2005-2014, SD-WACCM actually matches the upper stratospheric MLS variability fairly well, although these variability values are small. There have been difficulties in fully understanding (or modeling) observed upper stratospheric HCl variations before the declining phase that started after about 2000 (Waugh et al., 2001); see also Sect. 5.

~~To complete this discussion of variability, W~~We also show the ratios of model to data variability for stratospheric HNO_3 (2005-2014) and N_2O (2005-2012) in Fig. S13~~22~~. We already discussed the issues with missing model chemistry for upper stratospheric HNO_3 variability, as well as the low signal-to-noise issue for HNO_3 data at low latitudes in this region ~~(see bottom two panels, showing low variability ratios there)~~. There is reasonably good agreement in the HNO_3 variability between SD-WACCM and MLS for the lower to mid-stratosphere, while FR-WACCM generally overestimates the HNO_3 variability in this region. Fig. S13~~22~~ shows N_2O results ~~extending~~ down to 100 hPa. Here, ~~the~~ MLS N2O-640 data (from the 640 GHz radiometer) for 2005-2012 are used; these retrievals were curtailed in the first half of 2013 as a result of degradation in the 640 GHz radiometer signal chain. Based on results shown in SPARC (2017), there appears to be good agreement in the tropical interannual variability comparisons for N_2O at 100 hPa between MLS and other satellite-derived results. The lower stratospheric N_2O time series behave more smoothly at low latitudes in the models than in the observations. The interannual variability in the MLS N_2O measurements there is somewhat smaller than the standard deviations in monthly mean N_2O values (of ~~order~~ 20-30 ppbv). The MLS N_2O measurement noise itself for a monthly zonal mean (made up of about 5000-6000 profiles) should be less than 1 ppbv. Smoothing the model in the vertical domain to better match the MLS vertical resolution would not lead to a better fit to the observed variability. However, we should keep in mind that the MLS-derived N_2O variability is a small percentage (< 3%) of the monthly zonal mean N_2O abundances. In summary, SD-WACCM shows some underestimate of the observed lower stratospheric tropical variability for all the species considered here, except for HNO_3; FR-WACCM does so also for 3 species (H_2O, HCl, and N_2O). It may be that some of the larger variability in the measurements arises from effects not tied just to MLS radiance noise issues, or from variability caused by the proximity to the tropopause for measurements with finite vertical resolution; WACCM could also be genuinely underestimating the actual atmospheric variability near the tropopause (for unknown reasons).

5.2 Trends

In this section, we discuss how the WACCM runs compare to stratospheric observations when it comes to trends, from fairly short-term trends (from the Aura MLS time period) to longer-term trends based on comparisons with O_3, H_2O, and HCl GOZCARDS data records (see Sect. 2 and Froidevaux et al., 2015). Trend analyses have their own complexities in terms of analysis methods and uncertainty estimates. For example, O_3 trend assessments have had to deal with trend estimates from different long-term data records, each with its own characteristics (Tummon et al., 2015; WMO, 2014; Harris et al., 2015; Steinbrecht et al., 2017; Ball et al., 2017, 2018). This kind of analysis is especially difficult when investigating trends from time series with high variability compared to the size of the change over time, which is certainly an issue for the lower stratosphere~~ in particular~~. Also, g~~G~~lobal modeling efforts have ~~also~~ led to improved characterizations of the expected ~~long-term impact of different forcings, including the~~ combined and separate impacts on ozone profiles of long-term changes in halogen source gases and greenhouse gases (WMO, 2014).

Here, we focus mostly on ~~whether we obtain significant differences between~~ trend results from WACCM and ~~from~~ observations, given the application of the same analysis methods for the different~~three sets of time~~ series~~, for ozone and for other species~~. We have applied multiple (or multivariate) linear regression (MLR) to the time series of deseasonalized anomalies from the data, FR-WACCM, and SD-WACCM. In the Appendix (A3), we provide more details regarding the regression model, which includes commonly used additive functional terms, namely a linear trend and a constant term, cosine and sine functions with annual and semi-annual periodicities, as well as functions describing well known variations arising from the QBO and the El Niñ~~n~~o southern oscillation (ENSO~~);)~~. ~~the same functions are applied to fit the model anomaly time series as well.~~ Examples of observational time series from merged ozone observations for 1998 through 2014 are provided in Fig. A3, along with the fits to the series and the linear components (trends). Given the use of fairly short-term time series here (~~e.g.,~~ Aura MLS data alone), we have not included a solar cycle component in the fits, as it can be highly correlated with a linear trend, and more than one 11-yr cycle would be useful to better enable a separation of this signal. We also discuss our methodology for trend error evaluations in the Appendix (A3). We use a block bootstrap method, like the approach of Bourassa et al. (2014) for their trend analyses of ozone from the OSIRIS retrievals. We use random resampling of the residuals in yearly blocks (with 20,000 samples) in this Monte Carlo approach to estimating errors, and we display the trend error bars as 2σ values (which is very close to the 95% bounds on the distribution of linear trend results). Such calculations often lead to significantly larger error bars than more standard methods, which neglect the autocorrelation of residuals. ~~Again, we use the same regression model fits to extract trends from both WACCM and observed time series; error bar calculations are also applied the same way to all the time series.~~

For ozone, we give an overview in Fig. 18~~23~~ of percent deseasonalized anomaly time series ~~(expressed as a percent of long-term means)~~ for 3 latitude bins (northern mid-latitudes, tropics, and southern mid-latitudes) and 2 pressure levels (3.2 hPa for upper stratosphere, 68 hPa for lower stratosphere). The series were ~~first~~ deseasonalized in 10° latitude bins and then averaged. The GOZCARDS data ~~record~~ used here (version 2.20) is an update to the original (version 1.01) record (Froidevaux et al., 2015), as mentioned in Sect. 2. Fig. 18~~23~~ shows generally good agreement between the various time series, although if one looks carefully, SD-WACCM is generally closer to the observational time series than FR-WACCM is, as one might expect from previous considerations of goodness of fit and variability; also, percent variability is larger in the lower stratosphere than in the upper stratosphere~~ (note the y-axis range difference for these regions)~~, thus rendering trend detection more difficult at lower altitudes. We compare in Fig. 19~~24~~ the ozone profile trend results from MLS data alone, from ~~the period~~ 2005 through 2014, to those from FR-WACCM and SD-WACCM, for the 3 aforementioned latitude bins. We show the error bars as 2σ estimates, as we find this to be an easy way to visualize if there are ~~really~~ significant differences between models and data, or ~~indeed,~~ between ~~the~~ models. Figure 19~~24~~ provides~~ shows that there is~~ a robust indication from the MLS data that ~~the~~ upper stratospheric ozone values have been on the upswing in the past decade, at a rate of ~~about~~ 0.2 to 0.4%/yr, depending on latitude region, with 2σ uncertainties of ~0.2%/yr. While the 2σ error bars (obtained using the bootstrap method mentioned earlier) are fairly large, there are several ~~independent~~ latitude regions and pressure levels with similar results, and ~~thus,~~ this positive trend is a robust near-global upper stratospheric result. These results are broadly consistent with O_3 trends obtained by Steinbrecht et al. (2017), who use MLS as part of the longer-term merged data records, although they studied a longer time period (2000-2016). All things being equal, the errors in these trends should diminish as more years of data are added to the MLS O_3 record, which, for the middle and upper stratosphere, has been characterized as "very stable", namely within 0.1 to 0.2%/yr versus sonde and lidar network ozone data (Hubert et al., 2016); it ~~currently~~ seems difficult to quantify "absolute stability" to much better than this, especially in the lower stratosphere. In the lower stratosphere, trend results are~~tend to be~~ closer to zero, with larger variability and error bars (in %/yr), and unambiguous detection of post-1997 ozone trends in this region remains elusive (WMO, 2014; Harris et al., 2015). The 2005-2014 trends in Fig. 19~~24~~ show good broad agreement between model and data, with a tendency for SD-WACCM to agree better than FR-WACCM with MLS, albeit not significantly so, given the size of the error bars, especially in the lower stratosphere.~~;~~ T~~t~~he lower stratospheric tropical results from FR-WACCM are negative, in contrast to both the observation~~sal result~~ and~~s well as~~ SD-WACCM, but with large overlapping error bars~~ (given the variability and fairly short time period)~~. In the tropical upper stratosphere, both models exhibit a somewhat more positive trend than observed for this period, although, again, these trend differences are not statistically significant. Nevertheless, we find it rather striking that, as a function of latitude (in 10°-wide bins) and pressure, the 2005-2014 SD-WACCM O_3 trends ~~do~~ follow the MLS trend results quite well; this is clearly shown in Fig. 20,~~25~~ for central latitudes from 55°S to 55°N. The agreement in these patterns is not quite as good for FR-WACCM (not shown in Fig. 20); this is mostly evident in the tropical lower stratosphere (see Fig. 19).

For a consideration of longer time periods, we compare~~use~~ in Fig. 21 trends from the merged O_3 ~~data record from~~ GOZCARDS record (version 2.20) to ~~display~~ NH mid-latitude and tropical model trends ~~from the models~~ for 3 different periods ~~in Fig. 26~~. The top panels for 1985-1997 focus on the main "declining phase", while the middle panels (1998-2014) show ~~results for~~ the (expected) "early recovery" stage; the 2005-2014 (bottom panels) results are basically ~~essentially the same trends as~~ those from Aura MLS (Fig. 19~~24~~)~~ for that period~~. The largest differences between the two GOZCARDS data versions occur in the tropical upper stratosphere for the declining ozone phase; Fig. S14~~9~~ displays the tropical trend differences that we obtain for the same three periods as in Fig. ~~2~~19~~4~~. In agreement with this are the trend differences provided by Ball et al. (2017), who showed results for the original (version 1.01) GOZCARDS data and for SWOOSH. GOZCARDS version 2.20 data are now in better agreement with the merged SWOOSH O_3 product (as both use SAGE II version 7 data); also, Steinbrecht et al. (2017) showed that these two merged records lead to similar (post-2000) trend results. ~~As mentioned in Sect. 2,~~ T~~t~~he improvements in GOZCARDS version 2.20 ozone are a result of the incorporation of the SAGE II v7 retrievals (Damadeo et al., 2013), and the use of the MERRA temperatures (used in v7) for the conversion from density/altitude to the GOZCARDS mixing ratio/pressure grid. We note, however, that the lower stratospheric region exhibits interannual variability that is several times larger than that in the upper stratosphere, as seen in Fig. A3 for tropical 1998-2014 data versus SD-WACCM~~ anomaly time series~~. Even fairly subtle differences in time series over a few years can lead to a sign change in the trends, although there is no statistical significance in the resulting trend differences (see Fig. A3 and Fig. 21~~6~~, middle right panel at 68 hPa, for the data and SD-WACCM trend results).

To put in perspective what a statistically significant trend difference looks like, we show in Fig. S15~~0~~ the O_3 anomaly~~ time~~ series for 1998-2014 at 1 hPa for 30°N-60°N, where the SD-WACCM and GOZCARDS trend results lie outside their respective 2σ error~~uncertainties~~s (based on Fig. 21~~6, middle left panel~~); here~~in this case~~, ~~the ~~FR-WACCM results happen to be in better agreement with the data. One aspect that could impact model/data differences is that the models are not sampled, here, following the sparser (occultation) viewing, neither in latitude nor in time (time within each month and local time also, since model values are~~represent~~ 24-hr averages). D~~Much d~~enser spatial and temporal sampling is obtained during the MLS ~~observation ~~period, with very regular sampling; while small systematic (sampling-related) differences may affect comparisons between ~~average ~~models and MLS~~ abundances~~, such differences should be very consistent from year to year, thus minimizing the impact on ~~derived ~~trend differences. Also, some of the differences in the upper stratosphere might arise because the averaging of sunset and sunrise occultation data is not as robust for 1998-2004 as for pre-1998, when SAGE II was operating more continuously in both sunset and sunrise modes (and O_3 varies more strongly with local time at 1 hPa than at lower altitudes). Also, HALOE had decreasing spatio-temporal coverage in ~~the ~~later years;~~, but~~ thus, these upper stratospheric trend differences c~~w~~ould require more~~further~~ detailed investigations. For the 30°S-60°S region (not shown here), we observe similar results ~~(within error bars) ~~as in Fig. 21~~6~~ for 30°N-60°N; this includes small negative trends in the lower stratosphere for all 3 time periods, although with no statistical significance (i.e., consistent with a zero or slightly positive trend). We also see the same sort of vertical shift in the SD-WACCM profile trends for 1985-1997, as compared to the observed trend profile, which reaches its most negative value at a slightly higher altitude~~ (lower pressure)~~, but the reason for this shift is unknown; ~~theoretically, ~~this could be tied to model/data differences in the ClO peak~~,~~ and its evolution. Our ozone trend results are largely consistent with other previous work (references mentioned above), which (for records including MLS) typically used merged O_3 from GOZCARDS or from SWOOSH (Davis et al., 2016). We find statistically significant trends (meaning that~~, assuming Gaussian statistics,~~ a zero trend is not included in the 2σ error bar range) mostly in the upper stratosphere, both pre-1997 and after 1998. While we observe some small O_3 decreases in the lower stratosphere post-1998, as obtained recently also by the novel analyses of Ball et al. (2018), our study finds little statistical significance there, and a fair level of sensitivity to the starting year or to the data sets used, with a swing to more positive (but marginally significant) results, if the starting year is 2005~~ and one just uses MLS data~~. Past work (e.g., Harris et al., 2015) has also shown sensitivity to ~~the series ~~start~~ing~~ and end points; also, different regression analysis methods can ~~also ~~lead to ~~some ~~non-negligible differences (e.g., ~~, as shown for example by ~~Nair et al., ~~(~~2013;~~)~~ ~~and ~~Kuttippurath et al., ~~(~~2015). We also note that past analyses of lower stratospheric tropical O_3 data have shown positive tendencies, based not just on satellite data as indicated here (with marginal significance)~~ from MLS data alone~~, but also based on ~~earlier ~~SCIAMACHY data (Gebhardt et al., 2014); in that work~~is reference~~, a positive trend was also seen~~obtained~~ in~~from~~ averaged tropical ozonesonde data. This will continue to require further study, towards a longer-term result.

For H_2O, there have been ~~somewhat ~~conflicting past results on stratospheric trends, depending on whether one investigates sonde or satellite data (e.g., Oltmans et al., 2000; Scherer et al., 2008), and regarding mechanisms that could account for more than a few tenths of a %/yr increase in H_2O, as CH_4 increases do not appear to be large enough for this. ~~Beyond the potential impact of CH_4 trends on upper stratospheric H_2O, ~~C~~c~~hanges in cold point temperatures or in the circulation need to be invoked in order to account for significant decadal-scale trends in H_2O (e.g., Randel et al., 2000; Rohs et al., 2006; Tian and Chipperfield, 2006; Hegglin et al., 2014). Based on our analyses for 2005-2014~~ (MLS data versus models)~~, we find in Fig. 22~~7~~ that this recent decade shows a positive H_2O trend both in the MLS data and in ~~the ~~SD-WACCM~~ result~~, which tracks ~~the ~~observations (versus latitude as well as pressure) better than FR-WACCM does~~, as already seen in terms of quality of fits and variability~~. FR-WACCM exhibits systematically smaller H_2O trend values than both MLS and SD-WACCM at all pressures except near 100 hPa, although the FR-WACCM and SD-WACCM trend error bars (2σ) overlap. This overlap for FR-WACCM and MLS is more marginal in the lower mesosphere, where the impact on H_2O from CH_4 decomposition should be at its maximum, and cold point temperature variability issues are smaller than near the tropical tropopause. **An analysis of H_2O HALOE profiles and ground-based microwave profiles over Hawaii (Nedoluha et al., 2009) showed that changes in upper stratospheric and mesospheric H_2O are sensitive to the solar cycle (see also the mesospheric GOZCARDS H_2O time series in Froidevaux et al., 2015), but show only negligible overall trends between 1992 and 2008.** We can also see this evidence for smaller trends in the earlier portion of the H_2O record in time series (not shown here) from the GOZCARDS data set, with the positive part of the trend (as shown in the Fig. 22 trend results) coming after 2007. The cause for CH_4 changes over the past few decades have been difficult to identify with confidence (see, Feldman et al., 2018; Turner et al., 2017). Simple algebra indicates that such CH_4 changes can lead to only part (about half) of the H_2O increases reflected in the Fig. 22 trends, and there must be other reasons for these fairly large short-term trends. We believe that the significant decadal variability in H_2O, which arises from cold point temperature variability, propagated upward as a "tape recorder" signal, as well as QBO variability, account for a large part of the large positive H_2O trends over 2005-2014. Indeed, Garcia et al. (2007) noted in their studies of WACCM trends that multiple decades are likely needed to enable detection of the underlying secular rise in stratospheric H_2O, given the variability arising from ENSO, cold point temperature changes, the QBO and other factors. Also, a sudden drop in water vapor after 2000 can lead to a stronger post-2000 stratospheric H_2O trend, if one is considering a rather short time period; in our view, this plays a role in the large H_2O trends of Fig. 22. The fact that FR-WACCM trend results for the 2005-2014 period (see Fig. 22) are significantly smaller for the mid-stratosphere to lower mesosphere than the SD-WACCM and observed trends appears to mainly be a result of slightly different decadal variability in this run; we also see instances in longer time series (not shown) where FR-WACCM short-term changes appear to be larger than those from SD-WACCM. Finally~~Also~~, the FR-WACCM trends have smaller error bars, given the lower variability ~~typically~~ found in this model over the time period investigated here. With such lower variability, detection of a given trend would take less time than with the actual (observed) variability.~~; however, if the trend is larger (as shown by SD-WACCM or MLS), it also becomes easier to detect.~~

We should also note that ~~some~~ drifts have been detected between coincident MLS and sonde H_2O data, mostly since about 2010, with ~~implying that~~ MLS-derived trends being~~are~~ more positive than those from frost point profiles ~~at several sites~~ (Hurst et al., 2016). This relative drift (of as much as 0.5-1.6%/yr for 2010-2015, depending on altitude and location) could ~~therefore~~ play a role in the (small, in comparison) discrepancy between model (SD-WACCM) and MLS trends. ~~The~~ SD-WACCM results agree ~~quite~~ well with MLS in Fig. 22~~7~~, but they would become larger than MLS (adjusted) results if one were to subtract more than 0.1-0.2%/yr from these MLS trends. Possible causes for~~of observed~~ drifts between MLS and sonde H_2O data are ~~still~~ being investigated, with only a small part of this discrepancy currently attributable to a known instrument~~al~~ degradation issue for MLS H_2O, which probably ~~also~~ impacts other MLS data from the 190 GHz spectral region (notably N_2O~~ in particular~~). We note from Fig. 22~~7~~ that H_2O trend~~ value~~s from ~~both~~ models and MLS ~~data~~ agree better at 100 hPa. Also, we see that tropical H_2O ~~in the tropical region~~ at 80 to 100 hPa (near the stratospheric water vapor entry level) does not display much of a trend. S~~Other recent s~~tudies of entry-level H_2O using ~~large-scale~~ satellite data ~~and longer-term analyses (starting in the mid-1980s)~~ have concluded that no significant long-term trend is discernible (Hegglin et al., 2014; Dessler et al., 2014). The former study led to slightly negative lower stratospheric H_2O trends (although with no statistical significance).

For the 1992-2014 period and using GOZCARDS H_2O data (not shown here), we also find negative, although not statistically significant, trends in the lower stratosphere, with small (< 0.2%/yr) positive trends in the upper stratosphere and lower mesosphere.~~;~~ ~~T~~there is close agreement (within ~0.1%/yr) between ~~the~~ FR-WACCM and SD-WACCM trends ~~(typically within 0.1%/yr)~~ for that period. ~~This~~ GOZCARDS H_2O~~record~~, however, does not include SAGE II data back to the late 1980s, as was the case in the work~~for the analysis~~ by Hegglin et al. (2014), who also obtained positive long~~er~~-term (1980-2010) trends in the upper stratosphere from satellite-derived H_2O anomal~~ies~~~~y records~~, merged using a global CCM as a transfer function. As found by others, especially when dealing with ~~relatively~~ large decadal-type variability ~~(including the QBO)~~, the choice of start and end points, as well as the length of period studied, can significantly influence ~~the~~ trend values, ~~whether it~~ be it for H_2O or ~~for~~ O_3. ~~Long-term L~~Lower stratospheric H_2O trend detection is rendered difficult by such variability, including significant short-term ~~water vapor~~ changes (Randel et al., 2006; Hurst et al., 2011; Fueglistaler, 2012; Urban et al., 2014), as noted here also (~~(~~for ~~the relatively short time series of~~ Fig. S12)~~8)~~.

For HCl, ~~the~~ changes in stratospheric values~~abundances~~ have been non-linear, with a rapid rise prior to 1998, and a slower ~~rate of~~ decrease after 2004, as expected from ~~the~~ time-shifted abundances of total surface chlorine ~~at the surface~~ (Froidevaux et al., 2015). Focusing first on 2005-2014, we show the corresponding model and data HCl tren~~d~~d~~ result~~s for the lower stratosphere in Fig. 23~~8~~. The agreement between ~~the~~ SD-WACCM and MLS trends is ~~quite~~ good, especially for ~~the~~ 30°S-60°S ~~bin~~, although the error bars are fairly large. However, the (negative) HCl trend results from both models nearly~~practically~~ always lie below the observed trends. ~~for the three latitude bins shown in Fig. 28.~~ The upper portion of this model/data bias follows what we observe also~~s well~~ in the upper stratosphere (not shown), where MLS-derived HCl trends are ~~clearly~~ too ~~flat (~~shallow~~)~~ compared to expectations (from model and surface-derived chlorine trends), whereas ~~the~~ upper stratospheric ~~(negative)~~ trends from the original MLS HCl product were more negative (see Froidevaux et al., 2006; Livesey at al., 2018). As a reminder, the MLS team recommends that ~~ation is for~~ data users not ~~to~~ include upper stratospheric MLS HCl data (post-2006) in any trend studies. For the lower stratosphere, where the HCl line is broader, there is less concern about the inability to track the HCl trend.~~;~~ A~~a~~lso, the near-zero drifts (~~i.e.,~~ drifts < 0.1%/yr) obtained between two separate MLS O_3~~ozone~~ band retrievals (not shown ~~here~~), one from the same radiometer as HCl, and one from the main (very stable) standard MLS O_3~~ozone~~ product (see Hubert et al., 2016), provide some confidence regarding the stability of lower stratospheric HCl trends. At low latitudes, MLS HCl shows a positive trend (largest and statistically significant at 68 hPa, per Fig. 23~~8~~). The vertical gradient ~~structure~~ in these ~~observed HCl~~ trends is duplicated to some extent by ~~the~~ SD-WACCM results, although the model trends are always less than those ~~derived~~ from MLS. Latitude/pressure trend variability, including positive tendencies, could be related to circulation changes, as implied by analyses of short-term increases in lower stratospheric HCl seen in both ground-based and GOZCARDS data (Mahieu et al., 2014). Given the rapid rise in chlorine prior to 1998 and the non-linear changes near the peak period, we show in Fig. 24~~9~~ ~~some of the~~ lower stratospheric time series ~~(for 3 latitude bins and 3 pressures)~~ from GOZCARDS merged HCl (Froidevaux et al., 2015) and ~~the~~ WACCM ~~runs~~ for 1992-2014. There is fairly good agreement in the non-linear behavior observed in both data and model ~~lower stratospheric~~ series. The scatter in HCl data decreases after 2005, and the earlier time series suffer from more inhomogeneous sampling, which may at least in part explain the larger scatter and model/data differences (there is no attempt here to sample ~~the~~ models within each month like the data, and this would be difficult for a merged data set). There are also regions and periods of slow HCl increases ~~in HCl~~ in both data and models (Fig. 24~~9~~), as well as hemispheric differences in ~~the~~ short-term tendencies (~~;~~ ~~as discussed before by~~ Mahieu et al., ~~(~~2014;~~)~~ ~~and~~ Froidevaux et al., ~~(~~2015). The HCl time series are tracked ~~fairly~~ well by SD-WACCM, which generally matches the data better than FR-WACCM; this is consistent with the understanding that dynamically-driven variations are better captured by the incorporation in SD-WACCM of realistic meteorological fields (MERRA). Stolarski et al. (2018) ~~have recently~~ investigated the removal of dynamical variability from MLS lower stratospheric HCl ~~time~~ series by using MLS N_2O data~~measurements~~ as a fitting function~~parameter~~ in the regression ~~analysis~~; this led to ~~retrieved~~ HCl trends that ~~that generally~~ match expectations based on ~~the rates of~~ chang~~decrease~~es in surface ~~total~~ chlorine. The search for detailed explanations of such short-term increases and variability in lower stratospheric HCl (and other ~~changes in~~ composition changes) continues to be an interesting area of investigation.

In the upper stratosphere, it has been difficult to explain the details of ~~the~~ observed HCl variations from~~between~~ 1998 to~~and~~ 2002, including the dip between these years (Waugh et al., 2001). We show in Fig. S16~~1~~ ~~the~~ near-global (60°S-60°N) GOZCARDS HCl ~~time~~ series ~~from GOZCARDS~~ at 1 hPa. This shows~~helps to underscore that there is~~ a systematic model underestimate of HCl in the uppermost stratosphere; the model/data difference is much smaller ~~if one moves to~~ at pressures closer to 5 hPa. While the systematic ~~data~~ uncertainties (of ~0.2 ppbv~~, based on Froidevaux et al., 2015~~) in the data encompass the model values, the model total should ~~actually~~ be increased by the chlorine contribution from very short-lived halogenated substances (VSLS~~s~~) to the stratosphere; although this contribution is only of order ~~believed to be less than~~ 0.1 ppbv (Carpenter et al., 2014), recent evidence suggests that there could be a somewhat larger stratospheric chlorine contribution from VSLS~~s~~ (Oram et al., 2017). Nevertheless, the historical maximum for total tropospheric chlorine ~~abundance~~ was about 3.65 ppbv (WMO, 2014), and this should be the maximum total chlorine expected in the uppermost stratosphere. ~~Upper stratospheric HCl should see a somewhat broader peak than at the surface, with a smaller and time-delayed maximum, depending on transport-related effects (and age of air spectrum, e.g., see WMO, 2010).~~ While WACCM includes the proper abundance and evolution of chlorine source gases ~~at the surface~~, maximum WACCM HCl in the upper stratosphere (and lower mesosphere) is just under 3.4 ppbv. It is also interesting that the gap between the models (both versions) and the data worsens from 1992 to 2000, with the HCl peak occurring later in the data (with a broader peak than in the models). After about 2002, the decrease in near-global HCl roughly follows the model decrease; additional years of HCl data from ACE-FTS should help refine this comparison. In terms of HCl trends, Hossaini et al. (2018) have recently shown that there are positive changes (by ~15%) in model upper stratospheric HCl trends since 2000, i.e. the HCl decreases are smaller, if one takes into account the likely impact of changes in stratospheric chlorine from VSLS.

For N_2O and HNO_3, lower stratospheric model trends are compared to the corresponding MLS data trends in Fig. 25~~30~~. We note that the MLS standard product right after launch was N2O-640 (retrieved from the 640 GHz radiometer ~~band~~ data), but it was discontinued after mid-2013, ~~as mentioned earlier,~~ as a result of a ~~rapid~~ hardware degradation issue affecting that band (N_2O only). The current MLS ~~standard~~ product, N2O-190, is retrieved from the 190 GHz band. Figure S17~~2~~ provides evidence of negative drifts in lower stratospheric N2O-190, apparently accelerating in the last few years, since the SD-WACCM and actual N_2O values would be expected to continue to rise slowly after the end date on this plot, notably in the tropical lower stratosphere, where N_2O should follow tropospheric trends. Indeed, tropospheric N_2O has been increasing at a fairly steady rate of ~0.26%/yr (WMO, 2014), consistent with the ~~underlying~~ model N_2O and MLS N2O-640 lower stratospheric increases at low latitudes (see Fig. S17~~2~~ and especially the tropical trends ~~obtained~~ in Fig. 25~~30~~ at 100 hPa). ~~The~~ FR-WACCM N_2O trends show slightly poorer agreement than SD-WACCM versus N2O-640, although this is not statistically significant. T~~he t~~ropical lower stratospheric MLS N2O-190 trend~~ result~~s (not shown~~ here~~) are negative (albeit with error bars that encompass small positive trends), but show some differences versus expectations and the N2O-640 results. As for HCl, interhemispheric differences in lower stratospheric N_2O trends are interesting in terms of their implications for effects relating to transport (age of air) and changes in the circulation. At lower pressure values, the N_2O trends do not mirror the tropospheric N_2O trends (in %/yr), and other factors play a role (~~age of air and~~ changes in circulation, QBO, ~~, as well as~~ N_2O photodissociation). ~~The asymmetric trend pattern between hemispheres, even if it is not a long-term trend, may well point primarily to short-term effects tied to asymmetries in the age of air, and therefore, in the circulation.~~ The asymmetry in age of air ~~results~~ obtained by analyses of (2002-2012) MIPAS SF_6 data (Haenel et al., 2015) could also be related to asymmetries in ~~the~~ N_2O tendencies. They found relatively older air in the northern extra-tropics and younger air in the southern extra-tropics; ~~. t~~T~~his could also imply opposite trends for N_2O between~~in the~~ southern and northern lower stratosphere. However, Bönisch et al. (2011) have pointed out that ~~different~~ tracers (e.g., ~~, like~~ O_3 and N_2O)~~,~~ can be impacted ~~in~~ differently ~~ways~~ by both vertical and quasi-horizontal transport effects, depending on their relative vertical and meridional gradients ~~in both vertical and meridional domains~~. Moreover, their work indicates that detailed attribution of tracer variations to structural changes in the Brewer-Dobson circulation is a complex matter, and short-term and longer-term changes may well have different characteristics. Our work ~~here~~ mainly identifies ~~some~~ similarities between some of the trend patterns versus pressure and latitude from SD-WACCM, in particular, and the observed trends, for certain ~~time~~ periods. For the HNO_3 trends (Fig. 25~~30~~), we also see good agreement between models and data for 2005-2014; latitudinal tendencies and interhemispheric differences therein are similar for model and data. The spatial gradients of these species are different in the lower stratosphere (HCl and HNO_3 increase with height, in contrast to N_2O), and we see that the decreasing HCl trends for 2005-2014 at 30°S-60°S (Fig. 23~~8~~), in particular, are qualitatively similar to those from HNO_3 in this region. For lower stratospheric HNO_3, there is an underlying trend part caused by the slow increases in N_2O, as we can observe in longer-term (1980 to present) model time series (not shown here). N_2O and H_2O (source gases for HNO_3) are significantly affected by the QBO and there is a strong related variability in lower stratospheric HNO_3. ~~Furthermore, substantial increases in stratospheric aerosols after large volcanic eruptions have influenced lower stratospheric HNO_3 via heterogeneous hydrolysis of N_2O_5 (Arnold et al., 1990; Rinsland et al., 1994), and this will impact HNO_3 trends that include volcanically-perturbed periods. We saw that seasonal enhancements in NOx coming down from the mesosphere can also affect HNO_3 at high latitudes.~~ Some observed short-term trend patterns in HCl, HNO_3, and N_2O are better captured by ~~the~~ SD-WACCM, ~~model~~ overall, than by FR-WACCM, as we show in Fig. S18~~31~~ for the 2005-2010 period, relevant to the results of~~from~~ Mahieu et al. (2014), who emphasized short-term HCl increases during this time. We note the correlation in these short-term trend results for HNO_3 and HCl, but an anti-correlation for N_2O versus HCl (and HNO_3).

~~Finally, to re-emphasize how difficult it can be to detect small underlying trends in the lower stratosphere, in particular, Fig. 32 shows deseasonalized model anomalies from SD-WACCM for the 25-year period between 1990 and the end of 2014, for HNO_3, as well as O_3, H_2O, and HCl; on this scale, N_2O variability in this region is much smaller (not shown). As noted previously, and from past work on this topic, fairly sharp drops in water vapor in this region occurred shortly after 2000 and 2011, but we also note the significant decadal-type variability in this region, besides the expected links to QBO- and ENSO-type variations. There are also non-negligible radiative implications surrounding such variations for H_2O (Solomon et al., 2010; Gilford et al., 2016; Wang et al., 2016); however, a slow underlying long-term evolution would take time to detect, given the variability. Moreover, the percent variability is even larger for other species, which correlate well with H_2O in this region; indeed, the correlation coefficients between these various time series are all between 0.64 and 0.86, although this multi-species agreement is poorer at many other pressure levels or latitudes. In Fig. S13, we show that the observed variability in this same region for O_3 and H_2O (for 1992-2014) is fairly well matched by the models, although this is somewhat less true for H_2O than for O_3. Besides the importance of circulation effects on tracers in this region, tropopause temperatures will also affect water vapor; this adds some complexity in terms of exactly modeling its variations in the tropical lower stratosphere.~~

6 Summary and discussion

C~~The c~~limatological averages from FR-WACCM and SD-WACCM for O_3, H_2O, HCl, N_2O, and HNO_3 generally compare favorably (within ~~the~~ 2σ estimated systematic errors) with ~~the~~ Aura MLS data averages for 2005-2014. Model ozone values are usually within ±5-10% of the ~~average~~ data~~observations~~, except in the UTLS. In the lowest ~~portion of the~~ stratosphere, SD-WACCM generally exceeds the observed ozone means by ~~about~~ 30-50%, with FR-WACCM showing a smaller overestimate; both models also ~~tend to~~ overestimate (by ~60%) the amplitude of the annual cycle in this region. Such differences require further investigations, but would appear to implicate (in part) a transport-related issue in the models. There is also a model low bias (by ~10-30%) from 215 to 261 hPa at low latitudes, which could largely be caused by a known MLS high ozone bias in this region. For H_2O, there is a model low bias (by 5-15%) versus MLS data in the upper stratosphere and most of the mesosphere, although some of this arises from a small high bias in MLS H_2O versus other satellite data~~ sets~~ (~~see ~~Hegglin et al., 2013). Also, the models significantly underestimate ~~the ~~average HNO_3 valu~~abundance~~es in the upper stratosphere, notably at high latitudes; this ~~largely appears to ~~stems from missing model ion chemistry, as it relates to particle precipitation effects in the mesosphere, followed by downward wintertime polar transport of enhanced NOx, and subsequent ~~seasonal ~~increases in HNO_3. There is also some model overestimation by SD-WACCM of MLS HNO_3 (by about 40%) at high latitudes for pressures larger than 100 hPa, although there is a need for further validation of ~~the ~~HNO_3 data in this region. In the lower stratosphere at high southern latitudes, ~~the ~~variations in polar winter/spring composition observed by MLS are generally well matched by SD-WACCM, the main exception being for the early winter rate of decrease in HCl, which is too slow in the model. Grooß et al. (2018) have provided some~~further~~ discussion of this discrepancy, which should have little impact on winter/spring polar ozone depletion; indeed, we find good ~~or better ~~agreement between ~~the ~~seasonal high latitude observations and SD-WACCM for O_3~~ozone~~ and other species.

Regarding the fitted variability tied to the AO and SAO, there are a few discrepancies between model-derived amplitude patterns and the corresponding MLS climatology features, but FR-WACCM and SD-WACCM appear to properly capture the primary processes governing these modes of variability. SD-WACCM generally matches the data sets slightly better than FR-WACCM does. The O_3 AO stratospheric amplitudes are within ~25% of the MLS AO amplitudes. For H_2O, both WACCM versions exhibit AO and SAO patterns that are generally consistent with the observations, and with recently published satellite-derived results (Lossow et al., 2017a); we also note the WACCM underestimation of H_2O AO and SAO amplitudes in the lower stratosphere, although this is the region with the smallest amplitudes (< 0.1 ppmv).

We have provided diagnostics for the fits between ~~the ~~WACCM runs and ~~the ~~MLS deseasonalized anomaly ~~time ~~series. These consist of the correlation coefficient (R^2 diagnostic) as well as a diagnostic of RMS differences (model versus data), divided by the RMS variability in the data; a combined diagnostic (the ratio of the above two diagnostics) is also used to help differentiate between the two model runs, which are often not too far apart. Not too surprisingly, SD-WACCM, which is driven by realistic dynamics versus time, generally matches the observed zonal monthly mean anomalies significantly better than does FR-WACCM. This holds for all five species that we considered, with larger values of R^2 and smaller values of the RMS difference diagnostic. In the tropical lower stratosphere, where there is ~~some ~~nudging to equatorial winds for FR-WACCM (and even more so for SD-WACCM, see Sect. 2), the FR-WACCM fits to the data are generally improved. However, some details of the observed ~~interplay between SAO, AO, and QBO ~~variations in tropical upper stratospheric ozone are better matched by SD-WACCM ~~variations ~~than by FR-WACCM. Also, FR-WACCM shows poorer agreement with observed seasonal polar winter/spring lower stratospheric variations than does SD-WACCM. Finally, in the mesosphere, the water vapor anomalies are better matched by SD-WACCM than by FR-WACCM.

Variability comparisons represent a more fair and useful metric in terms of the characterization of model quality, in particular for a free-running model. Thus~~o this end~~, we ~~have ~~compare~~d~~ the RMS interannual variability from the anomalies in both WACCM models and observations, using data from ~~; we have used both ~~MLS ~~data ~~(2005-2014) and ~~data ~~from longer-term ~~series based on ~~GOZCARDS ~~data records for ~~O_3, H_2O, and HCl series. One of the main features ~~from these comparisons ~~is that the H_2O variability from ~~the ~~lower stratosphere to ~~the ~~upper mesosphere is underestimated by both model~~ run~~s~~ used here~~; this underestimate can reach a factor of two, although more typically, it is of order 30%. This implies that a larger number of years would be needed to detect an actual trend in H_2O than if one uses a model-based prediction (from FR-WACCM); this number of years would be increased by a factor of 1.2 to 1.6, if one uses the two variability factors mentioned above. Apart from the

WACCM underestimate of observed H_2O variability, the observed lower stratospheric variations, including significant drops in ~~the~~ H_2O~~ abundance~~, are better tracked by SD-WACCM than by FR-WACCM. This also seems to have implications for the model/data trend comparisons. O~~The o~~zone variability is better represented by ~~the~~ WACCM~~ models~~, with model/data variability ratios typically within a factor of 0.8 to 1.2. Observed HCl variability is underestimated somewhat by FR-WACCM for ~~the~~ 1992-2003~~ period~~, but not for ~~the later (~~2005-2014~~) period~~; the sparser HALOE sampling, compared to MLS, could explain some of the underestimate for the early period, especially in ~~the~~ polar regions. For N_2O, there is also a model underestimate (from both FR-WACCM and SD-WACCM) of MLS-derived ~~the~~ lower stratospheric low latitude variability~~ observed by MLS~~, although this variability is a small percentage of the mean values.

Regarding trends, the model comparisons versus the longer-term ozone data record from GOZCARDS (version 2.20~~ being used here~~) show generally good qualitative agreement in the time series in different latitude bins for both upper and lower stratospheric change. It is clear from such ~~time~~ series that the larger percent variability in the lower stratosphere will continue to render trend detection in this region more difficult than in the upper stratosphere. Based on the Aura MLS O_3 data record itself, which has been deemed very stable (Hubert et al., 2016), we observe robust evidence, considering the 2σ error bars of ~0.2%/yr ~~(estimated using a block bootstrap method)~~, that there is a positive upper stratospheric O_3~~ozone~~ trend for 2005-2014, at a rate of ~0.2-0.4%/yr. This is true for all ~~three broad mid-latitude and tropical~~ latitude regions considered here (30°N-60°N, 30°S-60°S, and 20°S-20°N), although the evidence is more marginal for~~in the~~ SH mid-latitudes. The WACCM trends estimated using the same regression model as used for the MLS data (anomaly) series show generally good agreement with the data trends, although the error bars are fairly large~~ (for both data and models)~~. Furthermore, the observed trend ~~(relative)~~ dependence on latitude and pressure is ~~well~~ matched quite well by ~~the~~ SD-WACCM trend results. We have not considered the high latitudes in detail here~~in this work~~, in part because of the significant dynamical variability in that region. ~~In this regard,~~ Stone et al. (2018) recently analyzed model results at high latitudes in the upper stratosphere, and showed that the large variability in that region, including the effects of solar proton events, is likely to mask detection of recovery (for now), although autumn and winter should exhibit the strongest recovery signals. In the lower stratosphere, where larger variability exists, the trends we deduce from the data~~ sets~~ and models agree within fairly large error bars, but~~although~~ there is generally no statistical significance. While there is a tendency for the GOZCARDS merged O_3 record to show small decreasing trends for ~~the~~ 1998-2014, ~~the~~ trend results reverse to near-zero or slightly positive tendencies (albeit with no robust statistical significance) if one considers ~~the~~ MLS data alone ~~(for~~ 2005-2014~~)~~. SD-WACCM trend~~ result~~s seem to track these positive tendencies, although with not much~~robust~~ statistical significance~~, based on our analyses~~. The recent work by Ball et al. (2018) indicates a net O_3 decrease in the lower stratosphere from about 1998 to the recent few years; this does not contradict the possibility of a turn-around towards a more positive trend~~rate of change~~ in this region during the most~~re~~ recent decade~~10-12 years~~. The positive tendency noted here may get more robust through ~~with the~~ analyses of more years of high quality global ozone profiles, and possibly more aligned with longer-term model expectations. ~~Future detailed analyses of these issues with different regression models and other methods are certainly indicated.~~

For H_2O, the most statistically significant trend result is an upper stratospheric increase for the post-2005 time period, peaking at slightly more than 0.5%/yr in the lower mesosphere, with MLS and SD-WACCM results agreeing fairly well, and FR-WACCM showing significantly smaller increases. The larger discrepancies for FR-WACCM are likely to arise from its poorer correlations (than SD-WACCM) with cold point temperatures, as well as with QBO variations.~~.~~ As shown before by others, there are multiple factors that can influence low-frequency variability in H_2O; indeed, these recent short-term trends go beyond what on would expect from changes associated with a slow, secular increase in methane, even if some of the recent methane changes have been non-linear (Schaefer et al., 2016; Nisbet et al., 2016). Also, t~~T~~he fact that the last decade has seen more of an upper stratospheric and mesospheric H_2O increase than the previous decade appears to correlate with the very shallow maximum that occurred in the last cycle (number 24) of the solar flux, which seems tied to the shallower dip, and broader overall maximum, in upper mesospheric H_2O (see the H_2O and solar flux time series in Fig. 16 of Froidevaux et al., 2015). ~~However, the non-linear influence of recent changes in methane, which include a plateau from 1999-2006, with a return to rising abundances after that (Schaefer et al., 2016; Nisbet et al., 2016), would also need to be considered for the upper stratosphere.~~ There is also a caveat regarding MLS-derived H_2O trends, given the existence of ~~non-negligible~~ drifts between sonde and MLS H_2O data (Hurst et al., 2016), at least since about 2010. Such drifts can only be partly~~ially~~ explained by currently known instrumental degradation issues affecting the MLS retrievals of H_2O, with some impact on other data from the 190 GHz radiometer (N_2O, in particular). Thus, ~~the~~ MLS-derived H_2O trend~~ result~~s obtained here are likely to be upper limits; this could~~an probably~~ explain why ~~the~~ model H_2O trends (at least from SD-WACCM) currently lie on the low side of the observed trends. An upcoming update to the MLS retrievals should~~will~~ lead to a reduction (but not an elimination) of the aforementioned drifts between MLS and sonde H_2O data. There is a continued need for cross-comparison of the various (diminishing number of) satellite H_2O data sets, as well as H_2O profiles from satellites and sondes, hopefully leading to a better understanding and mitigation of instrumental issues and drifts between ~~different~~ water vapor observations.

Our HCl trend analyses reveal broad agreement between the lower stratospheric MLS data (2005-2014) and the models, but with some systematic differences. ~~As mentioned in the past, upper stratospheric MLS HCl data are not deemed to be reliable enough for trend studies, since the cessation (in 2006) of the primary target band retrievals for MLS HCl.~~ While decreases in global lower stratospheric HCl ~~decreases~~ are ~~generally~~ indicated for 2005-2014, there are some hemispheric differences, and a significant increase is suggested in the tropical data at 68 hPa, where ~~; however,~~ there is only a slight positive trend ~~there~~ from ~~the~~ SD-WACCM ~~result~~ (with no statistical significance). However, there have been past indications of short-term increases in lower stratospheric HCl (Mahieu et al., 2014). While ~~-~~the lower stratospheric vertical gradients of MLS HCl trends~~ results~~ ~~from MLS HCl~~ are duplicated to some extent by SD-WACCM, the model trends are always more negative~~on the low side~~. There is no ~~clear~~ preference for SD-WACCM or FR-WACCM ~~trend~~ results, based on the~~in~~ comparisons to ~~the~~ observed lower stratospheric HCl trends. ~~There have also been past indications of short-term increases in lower stratospheric HCl (Mahieu et al., 2014); the study of such short-term tendencies for implications regarding circulation changes are worth pursuing further (but outside the scope of this work).~~ We see~~There is~~ a need for more comparisons of the various HCl measurements, satellite-based and ground-based, as well as models, in order to better understand circulation influences on stratospheric composition, as well as potential measurement-related issues (e.g., from potential sampling differences or measurement~~potential~~ drifts). Part of the model/data systematic difference in HCl trends could be explained by the omission (in WACCM) of the impact of VSLS on stratospheric chlorine, as indicated by the work of Hossaini et al. (2018).

For N_2O, the asymmetry in MLS-derived trends (for 2005-2012) between hemispheres, with negative trends (of up to about -1%/yr) at~~in the~~ NH mid-latitudes and positive trends (of up to 3%/yr) at~~in the~~ SH mid-latitudes, is in agreement with the asymmetry that exists in SD-WACCM results. S~~The s~~mall observed positive trends of ~0.2%/yr in the 100 to 30 hPa tropical region are ~~also~~ consistent with model results (SD-WACCM in particular), which in turn are very close to the known rate of increase in tropospheric N_2O (at ~~a rate of~~ about +0.26%/yr, see WMO, 2014). In the case of HNO_3, the MLS-derived lower stratospheric trend differences (for 2005-2014) between hemispheres are opposite in sign to those from N_2O (whose spatial gradients are largely of a sign opposite to those from HNO_3) and in reasonable agreement with both WACCM results, despite large error bars compared to the size of the trends. More detailed analyses would be needed to try to relate such trend asymmetries to changes in age of air, or circulation, but the QBO is a large contributor to short-term trend results in the middle stratosphere, for these species and more generally.

Overall, the models and observations show good agreement in the trends, with somewhat better results for SD-WACCM, which displays good correlations in the trend behavior versus latitude and pressure. However, the error bars are non-negligible, and the choice of start and end dates can have a significant impact on trends or tendencies. Given the existence of significant short-term and decadal-type variations for several lower stratospheric species, one should be cautious not to assign, or extrapolate, a tendency based on even a decade of data, to an underlying longer-term trend.

The diagnostics provided in this WACCM model evaluation can help distinguish even fairly subtle differences between models and observations, as well as between models. The improved fits to observations from a specified dynamics model versus a free-running model are to be expected, but also need to be documented. We are also reminded that observations have their own systematic issues, and close collaboration between modeling groups and instrument teams can help untangle issues that might be ~~more driven by, or at least~~ influenced by~~,~~ species-dependent instrumental effects. <u>W</u>~~Especially w~~hen comparing longer-term model time series to observations, even small systematic effects such as measurement drifts, or data merging issues, can become important for trend diagnostics. Finally, independent CCMs are not created in the same ~~exact~~ way, and nudging approaches ~~for free-running and specified dynamics models~~ can vary; <u>also</u>, some models have an internally-generated QBO, but most do not. While this study focused on (CESM1) WACCM ~~runs~~, further studies of the differences between high quality observations and various ~~international~~ models of atmospheric composition would be useful, to put this work in perspective. This could also be expanded to include ~~some~~ species not considered in this work, and/or with more of a focus on the upper troposphere.

~~28~~

Appendix A

A1 Examples of model grades

A grading method that has been applied in some previous comparisons (e.g., see Douglass et al., 1999; Waugh and Eyring, 2008) between atmospheric model values (M_n) and observed values (O_n) utilizes Eq. (A1) below to arrive at grades between 0 and 1 (and if a grade is < 0, it can be set to 0):

$$grade = 1 - \sum_1^N \frac{| M_n - O_n |}{E_f \times \sigma_n} \quad \text{(A1)}$$

with index n (in a given time series) varying between 1 and N (the total number of monthly values being compared for a given latitude/pressure bin), and σ_n representing the error in the observations. While the error factor E_f should probably be set to 2 or 3, this gives grades that are too small (close to zero or negative) if one applies such a formula to the MLS O_3 or H_2O time series, specifically to data sets with pretty well defined total measurement errors (provided as 2σ error estimates, per Sect. 4, meaning an error factor of 2). The grades shown here in Figs. A1 and A2 correspond to error factors (E_f) of 2 and 4, respectively. Figure A2 leads to O_3 and H_2O grades that are more useful than Fig. A1; it also shows similarities with the diagnostic based on the RMS of the differences between model and data, as shown in Sect. 4 (see the description of this diagnostic in Appendix A2).

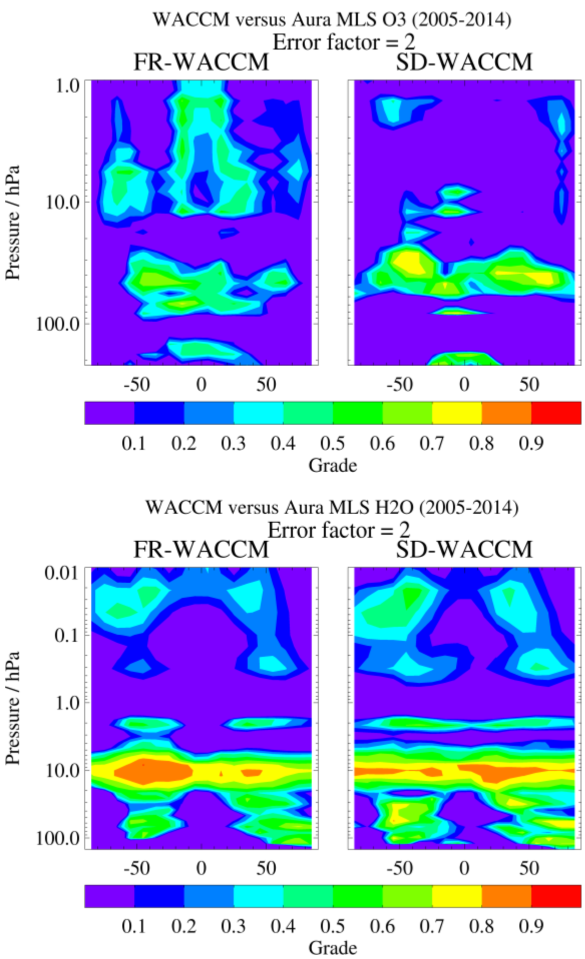

Figure A1. Examples of grades for model evaluations of O_3 (top two panels) and H_2O (bottom two panels), using a grading methodology that has been used in the past (see Eq. A1), applied to both FR-WACCM (left panels) and SD-WACCM (right panels) time series versus Aura MLS time series from 2005 through 2014. These grades are for an error factor of 2 (in Eq. A1).

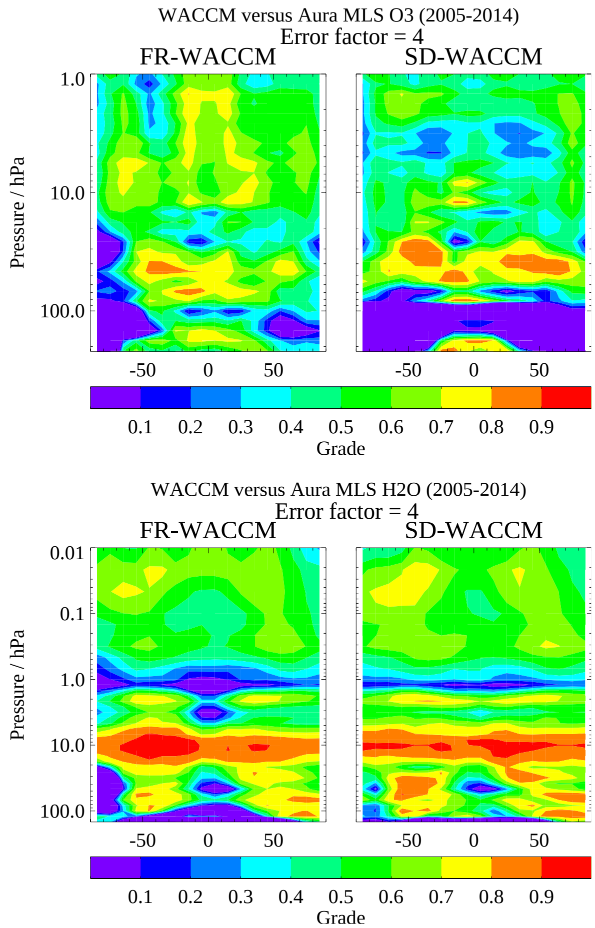

Figure A2. As in Fig. A1, for model evaluation grades of O_3 and H_2O (WACCM versus Aura MLS data), but for a value of 4 (rather than 2) for the error factor (see Eq. A1).

A2 RMS difference diagnostic

Given a model time series $M_i(t)$ and an observational time series (both series here representing deseasonalized anomalies) $O_i(t)$, the difference values between the two anomaly series are simply given by

$$\Delta_i(t) = M_i(t) - O_i(t) \qquad \text{(A2)}$$

and the root mean square (RMS) of these anomaly differences (RMSdif) is expressed as

$$RMSdif = \sqrt{\frac{1}{N} \sum_i \Delta_i^2} \qquad \text{(A3)}$$

The RMS value (variability) of the observational series (of anomalies) in this case is

$$RMS_O = \sqrt{\frac{1}{N} \sum_i O_i^2} \qquad \text{(A4)}$$

One of the diagnostics that we use in Sect. 4 to compare how well different models match up with the observed time series is given by

$$D_{RMSdif} = \frac{RMSdif}{RMS_O} \qquad \text{(A5)}$$

A3 Regression model

Functional form: The MLR model and fitting methodology used here is similar to the methods used by many others in the past, with a linear model description that uses annual, semi-annual, QBO, and ENSO terms. Thus, the model function to be fitted for coefficients a, b, c_n, d_n, f_1, f_2, and f_3 has the familiar form:

$$y(t) = a + b\,(t - t_m) + \sum_n (\, c_n \sin 2\pi t/P_n + d_n \cos 2\pi t/P_n\,) \\ + f_1\, QBO_1(t) + f_2\, QBO_2(t) + f_3 ENSO(t) \qquad \text{(A6)}$$

with the (monthly series) time variable expressed by t, and t_m chosen as the series mid-point; the linear trend term is coefficient b above. The *sine* and *cosine* functions provide for periodic variations with periods P_n. For our work, we use the two primary shorter-term periodic oscillations, annual (12-month period) and semi-annual (6-months), in Equation (A6). The QBO is also a major source of variability in stratospheric composition time series. As a QBO proxy, we include the variability in monthly mean tropical wind series; we use the linear combination of (roughly orthogonal) equatorial wind series at 50 hPa and 30 hPa as the QBO_1 and QBO_2 functions above, to account for phase shifts in the series at different locations. Monthly mean zonal equatorial wind data are made available by the Freie Universität Berlin (see http://www.geo.fu-berlin.de/en/met/ag/strat/produkte/qbo/ for data access information and references). We have also tested fits with the zonal mean wind vertical shear (gradient) rather than the wind itself as a proxy, but this did not make significant changes in the trends (or improvements in the residuals). The ENSO proxy follows the monthly mean multivariate ENSO Index (MEI), which combines data from six main geophysical variables over the tropical Pacific (see Wolter and Timlin, 1993, 1998; https://www.esrl.noaa.gov/psd/enso/mei/index.html). Also, since the solar cycle 11-yr term can be highly correlated with the linear term, especially for shorter-term records like the time series from MLS data alone, we have not added a proxy solar term in this study. For further discussions of alternate fitting methods (e.g., methods using effective equivalent chlorine time series as a proxy), the reader is referred to the abundant literature on trend assessments (see WMO, 2014 and references therein). Our main goal here is to retrieve trends and trend errors from the data and the models in the same way. An example of deseasonalized ozone time series in the tropics at two pressure levels is provided in Fig. A3, which shows MLS data and SD-WACCM time series, along with the fits and the linear trends.

Trend errors: For the evaluation of error bars in the linear trends, we have used the method of bootstrap resampling (Efron and Tibshirani, 1986). As others have done for ozone trend analyses (Randel and Thompson, 2011, Bourassa et al., 2014), we have applied this using block bootstrapping (using yearly blocks of data), thereby preserving some of the dependency in the time series. Basically, one samples and (randomly) replaces blocks of yearly data for a large number of resampling cases (on the residuals), and then calculates the standard deviation of the large number of trend results (linear fits) to arrive at the trend uncertainties; note that we use 2σ values as error bars in our comparisons (which is very close to 95% bounds). We have used 20,000 samples in our bootstrap analyses; changing this number (e.g., by several thousand) does not alter the results significantly, as long as one chooses a large enough total number of cases. An alternative method is to attempt to correct trend uncertainties for the autocorrelation of the residuals after the regression fit (Tiao et al., 1990; Weatherhead et al., 1998; Santer et al., 2000). The existence of non-random residuals effectively implies that the number of independent data points is less than the number making up the original time series. The end result is that trend uncertainties are larger than if one neglects these effects. We find that trend errors from this bootstrap method for our time series examples are often larger than more simplistic/standard calculations by factors ranging from about 1.2 to 2 or more. We have checked our trend error calculations with the OSIRIS team, based on a sample time series, as they have used the same block bootstrap approach (Bourassa et al., 2014).

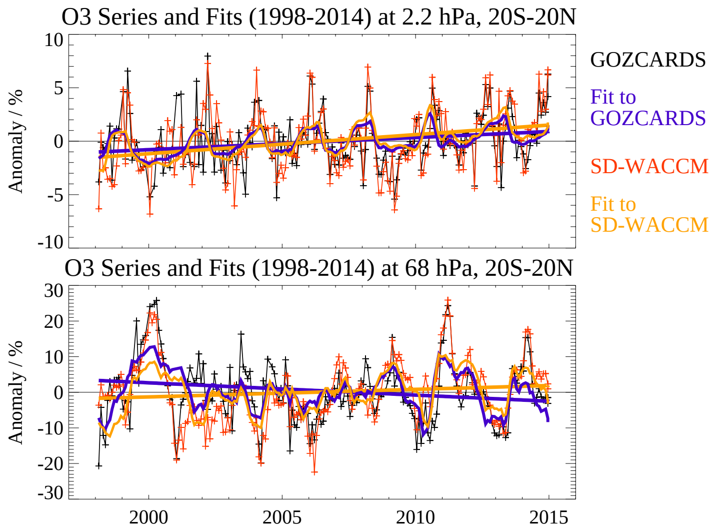

Figure A3. Deseasonalized monthly mean anomaly time series for O_3 (%) at 2.2 hPa (top panel) and 68 hPa (bottom panel), for 1998 through 2014, for averages over 20°S-20°N. Note that the y-axis range for 68 hPa (bottom) is 3 times larger than for 2.2 hPa (top). The data set (black) is from the GOZCARDS ozone record, with the SD-WACCM (simulated) series (red) also shown. Fits to the observational series are in purple, and fits to the model series are in orange; the fitted time series functions (curves) and the fitted linear components (straight lines) are shown.

Data availability. Aura MLS data used in this work are monthly zonal means derived from Level 2 MLS data, which are accessible from the Goddard Earth Sciences Data and Information Services Center (GES DISC), funded by NASA's Science Mission Directorate; a link for MLS Level 2 data access can be found at https://mls.jpl.nasa.gov/data/. GOZCARDS data sets can be obtained (by entering GOZCARDS in the search) at http://disc.gsfc.nasa.gov. More recent years (and version updates) will be made available at this site, or can be obtained by request to the first author. The WACCM model output used here is provided in some of the references and is available from the NCAR Earth System Grid at https://www.earthsystemgrid.org/search.html?Project=CCMI1. For WACCM, we thank NASA Goddard Space Flight Center for the MERRA data (accessed freely online at http://disc.sci.gsfc.nasa.gov/).

Author contributions. L. Froidevaux produced the majority of this manuscript, including the Figures. D. Kinnison provided the model runs used in these comparisons to observational data from Aura MLS and GOZCARDS; he also described the models used here and provided substantial guidance and comments for this manuscript. H.-J. Wang and J. Anderson, along with L. Froidevaux, were key participants in the development and creation of the GOZCARDS data sets, including the recently updated version (2.20) for merged ozone used herein. R. Fuller was another key participant in the GOZCARDS data production, and he also provided programming support for these model intercomparisons.

Competing interests. The authors declare that they have no conflicts of interest.

Special issue statement. This manuscript is intended as part of the Chemistry-Climate Modelling Initiative (CCMI) inter-journal special issue (ACP/AMT/ESSD/GMD SI).

Acknowledgements. Work at the Jet Propulsion Laboratory, California Institute of Technology, was performed under contract with the National Aeronautics and Space Administration (NASA). GOZCARDS data were initially produced under the NASA Making Earth System Data Records for Use in Research Environments (MEaSUREs) program, with data continuity/updates now provided under MLS funding. We acknowledge insights and comments regarding aspects of this work from Rolando Garcia, Nathaniel Livesey, Jessica Neu, and Michelle Santee. We also appreciate discussions and comparison work regarding trends and error bars with Doug Degenstein and Chris Roth, as well as Emmanuel Mahieu (and via some of his European contacts). The National Center for Atmospheric Research (NCAR) is sponsored by the U.S. National Science Foundation (NSF). WACCM is a component of NCAR's Community Earth System Model (CESM), which is supported by the NSF and the Office of Science of the U.S. Department of Energy. Computing resources for WACCM were provided by NCAR's Climate Simulation Laboratory, sponsored by NSF and other agencies; we also acknowledge the computational and storage resources of NCAR's Computational and Information Systems Laboratory (CISL).

References

Andersson, M. E., Verronen, P. T., Marsh, D. R., Pälvärinta, S.-M., and Plane, J. M. C.: WACCM-D-Improved modeling of nitric acid and active chlorine during energetic particle precipitation, J. Geophys. Res. Atmos., 121, 10,328–10,341, doi:10.1002/2015JD024173, 2016.

Ball, W. T., Alsing, J., Mortlock, D. J., Rozanov, E. V., Tummon, F., and Haigh, J. D.: Reconciling differences in stratospheric ozone composites, Atmos. Chem. Phys., 17, 12269–12302, https://doi.org/10.5194/acp-17-12269-2017, 2017.

Ball, W.T., Alsing, J., Mortlock, D. J., Staehelin, J., Haigh, J. D., Peter, T., Tummon, F., Stübi, R., Stenke, A., Anderson, J., Bourassa, A., Davis, S. M., Degenstein, D., Frith, S., Froidevaux, L., Roth, C., Sofieva, V., Wang, R., Wild, J., Yu, P., Ziemke, J. R., and Rozanov, E. V.: Continuous decline in lower stratospheric ozone offsets ozone layer recovery, Atmos. Chem. Phys., 18, 1379–1394, https://doi.org/10.5194/acp-18-1379-2018, 2018.

Bandoro, J., Solomon, S., Santer, B. D., Kinnison, D. E., and Mills, M. J.: Detectability of the impacts of ozone-depleting substances and greenhouse gases upon stratospheric ozone accounting for nonlinearities in historical forcings, Atmos. Chem. Phys., 18, 143-166, https://doi.org/10.5194/acp-18-143-2018, 2018.

Böhringer, H., Fahey, D.W., Fehsenfeld, F. C., and Ferguson, E. E.: The role of ion–molecule reactions in the conversion of $N2O_5$ to HNO_3 in the stratosphere, Planet. Space. Sci., 31, 185–191, 1983.

Bönisch, H., Engel, A., Birner, Th., Hoor, P., Tarasick, D. W., and Ray, E. A.: On the structural changes in the Brewer–Dobson circulation after 2000, Atmos. Chem. Phys., 11, 3937–3948, doi:10.5194/acp-11-3937-2011, 2011.

Bourassa, A. E., Degenstein, D. A., Randel, W. J., Zawodny, J. M., Kyrölä, E., McLinden, C. A., Sioris, C. E., and Roth, C. Z.: Trends in stratospheric ozone derived from merged SAGE II and Odin-OSIRIS satellite observations, Atmos. Chem. Phys., 14, 6983-6994, doi:10.5194/acp-14-6983-2014, 2014.

Brinkop, S., Dameris, M., Jockel, P., Garny, H., Lossow, S., and Stiller, G.: The millennium water vapour drop in chemistry-climate model simulations, Atmos. Chem. Phys., 16, 8125-8140, www.atmos-chem-phys.net/16/8125/2016, doi:10.5194/acp-16-8125-2016, 2016.

Calvo, N, Garcia, R. R., and Kinnison, D. E.: Revisiting Southern Hemisphere polar stratospheric temperature trends in WACCM: The role of dynamical forcing, Geophys. Res., Lett., 44, 3402-3410, doi:10.1002/2017GL072792, 2017.

Carpenter, L. J., Reimann, S., Burkholder, J. B., Clerbaux, C., Hall, B. D., Hossaini, R., Laube, J. C., and Yvon-Lewis, S. A.: Ozone-depleting substances (ODSs) and other gases of interest to the Montreal Protocol, chap. 1, in: Scientific Assessment of Ozone Depletion: 2014, Global Ozone Research and Monitoring Project – Report No. 55, World Meteorological Organization, Geneva, Switzerland, 2014.

Damadeo, R. P., Zawodny, J. M., Thomason, L. W., and Iyer, N.: SAGE version 7.0 algorithm: application to SAGE II, Atmos. Meas. Tech., 6, 3539–3561, https://doi.org/10.5194/amt-6-3539-2013, 2013.

Davis, S. M., Rosenlof, K. H., Hassler, B., Hurst, D. F., Read, W. G., Vömel, H., Selkirk, H., Fujiwara, M., and Damadeo, R.: The Stratospheric Water and Ozone Satellite Homogenized (SWOOSH) database: a long-term database for climate studies, Earth Syst. Sci. Data, 8, 461–490, https://doi.org/10.5194/essd-8-461-2016, 2016.

Dessler, A. E., Schoeberl, M. R., Wang, T., Davis, S. M., Rosenlof, K. H., and Vernier, J.-P.: Variations of stratospheric water vapor over the past three decades, J. Geophys. Res. Atmos., 119, 12,588-12,598, doi:10.1002/2014JD021712, 2014.

Dhomse, S., et al.: Estimates of Ozone Return Dates from Chemistry-Climate Model Initiative Simulations, Atmos. Chem. Phys., in review, 2018.

Douglass, A. R., Prather, M. J., Hall, T. M., S. E. Strahan, S. E., Rasch, P. J., Sparling, L. C., Coy, L., and Rodriguez, J. M.: Choosing meteorological input for the global modeling initiative assessment of high-speed aircraft, J. Geophys. Res., 104, 27,545-27,564, 1999.

Ern, M., Preusse, P., and Riese, M.: Driving of the SAO by gravity waves as observed from satellite, Ann. Geophys., 33, 483–504, www.ann-geophys.net/33/483/2015/, doi:10.5194/angeo-33-483-2015, 2015.

Eyring, V., et al.: Overview of IGAC/SPARC Chemistry-Climate Model Initiative (CCMI) Community Simulations in Support of Upcoming Ozone and Climate Assessments, SPARC Newsletter, 40, 48-66, January 2013, 2013.

Feldman, D. R., Collins, W. D., Biraud, S. C., Risser, M. D., Turner, D. D., Gero, P. J., Tadic, J., Helmig, D., Xie, S., Mlawer, E. J., Shippert, T. R., and Torn, M. S.: Observationally derived rise in methane surface forcing mediated by water vapour trends, Nat. Geosci., 11, 238-246, https://doi.org/10.1038/s41561-018-0085-9

Frith, S. M., Stolarski, R. S., Kramarova, N. A., and McPeters, R. D.: Estimating uncertainties in the SBUV Version 8.6 merged profile ozone dataset, Atmos. Chem. Phys., 17, 14695-14707, https://doi.org/10.5194/acp-17-14695-2017, 2017.

Froidevaux, L., Livesey, N. J., Read, W. G., Salawitch, R. J., Waters, J. W., Drouin, B., MacKenzie, I. A., Pumphrey, H. C., Bernath, P., Boone, C., Nassar, R., Montzka, S., Elkins, J., Cunnold, D., and Waugh, D.: Temporal decrease in upper atmospheric chlorine, Geophys. Res. Lett., 33, L23813, doi:10.1029/2006GL027600, 2006.

Froidevaux, L., Jiang, Y. B., Lambert, A., Livesey, N. J., Read, W. G., Waters, J. W., Fuller, R. A., Marcy, T. P., Popp, P. J., Gao, R. S., Fahey, D. W., Jucks, K. W., Stachnik, R. A., Toon, G. C., Christensen, L. E., Webster, C. R., Bernath, P. F., Boone, C. D., Walker, K. A., Pumphrey, H. C., Harwood, R. S., Manney, G. L., Schwartz, M. J., Daffer,W.H., Drouin, B. J., Cofield, R. E., Cuddy, D. T., Jarnot, R. F., Knosp, B. W., Perun, V. S., Snyder, W. V., Stek, P. C., Thurstans, R. P., and Wagner, P. A.: Validation of Aura Microwave Limb Sounder HCl measurements, J. Geophys. Res., 113, doi:10.1029/2007JD009025, D15S25, 2008a.

Froidevaux, L., Jiang, Y. B., Lambert, A., Livesey, N. J., Read, W. G., Waters, J. W., Browell, E. V., Hair, J. W., Avery, M. A., McGee, T. J., Twigg, L. W., Sumnicht, G. K., Jucks, K. W., Margitan, J. J., Sen, B., Stachnik, R. A., Toon, G. C., Bernath, P. F., Boone, C. D., Walker, K. A., Filipiak, M. J., Harwood, R. S., Fuller, R. A., Manney, G. L., Schwartz, M. J., Daffer,W.H., Drouin, B. J., Cofield, R. E., Cuddy, D. T., Jarnot, R. F., Knosp, B. W., Perun, V. S., Snyder, W. V., Stek, P. C., Thurstans, R. P., and Wagner, P. A.: Validation of Aura Microwave Limb Sounder stratospheric and mesospheric ozone measurements, J. Geophys. Res., 113, doi:10.1029/2007JD008771, D15S20, 2008b.

Froidevaux, L., Anderson, J., Wang, H.-J., Fuller, R. A., Schwartz, M. J., Santee, M. L., Livesey, N. J., Pumphrey, H. C., Bernath, P. F., Russell III, J. M., and McCormick, M. P.: Global Ozone Chemistry And Related trace gas Data records for the Stratosphere (GOZCARDS): methodology and sample results with a focus on HCl, H2O, and O3, Atmos. Chem. Phys., 15, 10471–10507, https://doi.org/10.5194/acp-15-10471-2015, 2015.

Fueglistaler, S.: Step-wise changes in stratospheric water vapor? J. Geophys. Res., 117, D13302, doi:10.1029/2012JD017582, 2012.

Fueglistaler, S. and Haynes, P. H.: Control of interannual and longer-term variability of stratospheric water vapor, J. Geophys. Res., 110, 24 108, D24 108, doi:10.1029/2005JD006019, 2005.

Funke, B., Baumgaertner, A., Calisto, M., Egorova, T., Jackman, C. H., Kieser, J., Krivolutsky, A., Lopez-Puertas, M., Marsh, D. R., Reddmann, T., Rozanov, E., Salmi, S.-M., Sinnhuber, M., Stiller, G. P., Verronen, P. T., Versick, S., von Clarmann, T., Vyushkova, T. Y., Wieters, N., and Wissing, J. M.: Composition changes after the “Halloween” solar proton event: The High-Energy Particle Precipitation in the Atmosphere (HEPPA) model versus MIPAS data intercomparison study, Atmos. Chem. Phys., 11(3), 9089–9139, doi:10.5194/acp-11-9089-2011, 2011.

Garcia, R. R., D. Marsh, D. E. Kinnison, B. Boville, and F. Sassi, Simulations of secular trends in the middle atmosphere, 1950-2003, J. Geophys. Res., 112, D09301, doi:10.1029/2006JD007485, 2007.

Garcia, R. R., Smith, A. K., Kinnison, D. E., de la Camara, A. and Murphy, D.: Modification of the gravity wave parameterization in the Whole Atmosphere Community Climate Model: Motivation and results, J. Atmos. Sci., 74, 275-291, doi:10.1175/JAS-D-16-0104.1, 2017.

Gebhardt, C., Rozanov, A., Hommel, R., Weber, M., Bovensmann, H., Burrows, J. P., Degenstein, D., Froidevaux, L., and Thompson, A. M.: Stratospheric ozone trends and variability as seen by SCIAMACHY from 2002 to 2012, Atmos. Chem. Phys., 14, 831–846, doi:10.5194/acp-14-831-2014, 2014.

Gille, J., Karol, S., D. Kinnison, D., Lamarque, J.-F., and Yudin, V.: The role of mid-latitude mixing barriers in creating the annual variation of total ozone in high northern latitudes, *J. Geophys. Res.*, 119, doi:10.1002/2013JD0214162014, 2014.

Grooß, J.-U., Müller, R., Spang, R., Tritscher, I., Wegner, T., Chipperfield, M. P., Feng, W., Kinnison, D. E., and Madronich, S.: On the discrepancy of HCl processing in the dark polar vortices, Atmos. Che. Phys. Discuss., https://doi.org/10.5194/acp-2018-202, 2018.

Haenel, F. J., Stiller, G. P., von Clarmann, T., Funke, B., Eckert, E., Glatthor, N., Grabowski, U., Kellmann, S., Kiefer, M., Linden, A., and Reddmann, T.: Reassessment of MIPAS age of air trends and variability, Atmos. Chem. Phys., 15, 13,161–13,176, doi:10.5194/acp-15-13161-2015, 2015.

Harris, N. R. P., Hassler, B., Tummon, F., Bodeker, G. E., Hubert, D., Petropavlovskikh, I., Steinbrecht, W., Anderson, J., Bhartia, P. K., Boone, C. D., Bourassa, A., Davis, S. M., Degenstein, D., Delcloo, A., Frith, S. M., Froidevaux, L., Godin-Beekmann, S., Jones, N., Kurylo, M. J., Kyrölä, E., Laine, M., Leblanc, S. T., Lambert, J.-C., Liley, B., Mahieu, E., Maycock, A., de Maziere, M., Parrish, A., Querel, R., Rosenlof, K. H., Roth, C., Sioris, C., Staehelin, J., Stolarski, R. S., Stubi, R., Tamminen, J., Vigouroux, C., Walker, K., Wang, H. J., Wild, J., and Zawodny, J. M.: Past changes in the vertical distribution of ozone - Part 3: Analysis and interpretation of trends, Atmos. Chem. Phys., 15, 9965-9982, https://doi.org/10.5194/acp-15-9965-2015, 2015.

Hegglin, M. I., Tegtmeier, S., Anderson, J., Froidevaux, L., Fuller, R., Funke, B., Jones, A., Lingenfelser, G., Lumpe, J., Pendlebury, D., Remsberg, E., Rozanov, A., Toohey, M., Urban, J., von Clarmann, T., Walker, K. A., Wang, R., and Weigel, K.: SPARC Data Initiative: Comparison of water vapor climatologies from international satellite limb sounders, J. Geophys. Res. Atmos., 118, 11,824–11,846, doi: 10.1002/jgrd.50752, 2013.

Hegglin, M. I., Plummer, D. A., Shepherd, T. G., Scinocca, J. F., Anderson, J., Froidevaux, L., Funke, B., Hurst, D. and Rozanov, A., Urban, J., von Clarmann, T., Walker, K. A., Wang, H. J., Tegtmeier, S., and Weigel, K.: Vertical structure of stratospheric water vapour trends derived from merged satellite data, Nat. Geosci., 7, 768–776, https://doi.org/10.1038/ngeo2236, 2014.

Holton, J. R., and Gettelman, A.:Horizontal transport and the dehydration of the stratosphere, Geophys. Res. Lett., 28, 2799–2802, 2001.

Hossaini, R., et al.: Recent trends in stratospheric chlorine from very short-lived substances, J. Geophys. Res., in press, 2018.

Hubert, D., Lambert, J.-C., Verhoelst, T., Granville, J., Keppens, A., Baray, J.-L., Cortesi, U., Degenstein, D. A., Froidevaux, L., Godin-Beekmann, S., Hoppel, K. W., Kyrölä, E., Leblanc, T., Lichtenberg, G., McElroy, C. T., Murtagh, D., Nakane, H., Russell III, J. M., Salvador, J., Smit, H. G. J., Stebel, K., Steinbrecht, W., Strawbridge, K. B., Stübi, R., Swart, D. P. J., Taha, G., Thompson, A. M., Urban, J., van Gijsel, J. A. E., von der Gathen, P., Walker, K. A., Wolfram, E., and Zawodny, J. M.: Ground-based assessment of the bias and long-term stability of fourteen limb and occultation ozone profile data records, Atmos. Meas. Tech., 9, 2497-2534, https://doi.org/10.5194/amt-9-2497-2016, 2016.

Hurst, D. F., Oltmans, S. J., Vomel, H., Rosenlof, K. H., Davis, S. M., Ray, E. A., Hall, E. G., and Jordan, A.: Stratospheric water vapor trends over Boulder, Colorado: Analysis of the 30 year Boulder record, J. Geophys. Res., 116, D02306, doi:10.1029/2010JD015065, 2011.

Hurst, D. F., Read, W. G., Vömel, H., Selkirk, H. B., Rosenlof, K. H., Davis, S. M., Hall, E. G., Jordan, A. F., and Oltmans, S. J.: Recent divergences in stratospheric water vapor measurements by frost point hygrometers and the Aura Microwave Limb Sounder, Atmos. Meas. Tech., 9, 4447–4457, www.atmos-meas-tech.net/9/4447/2016/, doi:10.5194/amt-9-4447-2016, 2016.

Imai, K., Manago, N., Mitsuda, C., Naito, Y., Nishimoto, E., Sakazaki, T., Fujiwara, M., Froidevaux, L., von Clarmann, T., Stiller, G. P., Murtagh, D. P., Rong, P.-P., Mlynczak, M. G., Walker, K. A., Kinnison, D. E., Akiyoshi, H., Nakamura, T., Miyasaka, T., Nishibori, T., Mizobuchi, S., Kikuchi, K.-I., Ozeki, H., Takahashi, C., Hayashi, H., Sano, T., Suzuki, M., Takayanagi, M., and Shiotani, M.: Validation of ozone data from the Superconducting Submillimeter-Wave Limb-Emission Sounder (SMILES), J. Geophys. Res.-Atmos., 118, 5750–5769, doi:10.1002/jgrd.50434, 2013.

Jackman, C. H., Marsh, D. R., Kinnison, D. E., Mertens, C. J., and Fleming, E. L.: Atmospheric changes caused by galactic cosmic rays over the period 1960-2010, *Atmos. Chem. Phys.*, doi:10.5194/acp-16-5853-2016, 2016.

Jensen, E. and Pfister, L.: Transport and freeze-drying in the tropical tropopause layer, J. Geophys. Res. Atmos., 109, D02207, http://dx.doi.org/10.1029/2003JD004022, 2004.

Jiang, Y. B., Froidevaux, L., Lambert, A., Livesey, N. J., Read, W. G., Waters, J. W., Bojkov, B., Leblanc, T., McDermid, I. S., Godin-Beekmann, S., Filipiak, M. J., Harwood, R. S., Fuller, R. A., Daffer, W. H., Drouin, B. J., Cofield, R. E., Cuddy, D. T., Jarnot, R. F., Knosp, B. W., Perun, V. S., Schwartz, M. J., Snyder, W. V., Stek, P. C., Thurstans, R. P., Wagner, P. A., Allaart, M., Andersen, S. B., Bodeker, G., Calpini, B., Claude, H., Coetzee, G., Davies, J., De Backer, H., Dier, H., Fujiwara, M., Johnson, B., Kelder, H., Leme, N. P., Koenig-Langlo, G., Kyro, E., Laneve, G., Fook, L. S., Merrill, J., Morris, G., Newchurch, M., Oltmans, S., Parrondo, M. C., Posny, F., Schmidlin, F., Skrivankova, P., Stubi, R., Tarasick, D., Thompson, A., Thouret, V., Viatte, P., Vömel, H., von Der Gathen, P., Yela, M., and Zablocki, G.: Validation of the Aura Microwave Limb Sounder Ozone by Ozonesonde and Lidar Measurements, J. Geophys. Res., 112, D24S34, doi:10.1029/2007JD008776, 2007.

Khosrawi, F., Mueller, R., Urban, J., Proffitt, M. H., Stiller, G., Kiefer, M., Lossow, S., Kinnison, D. E., Olschewski, F.,Riese, M., and Murtagh, D.: Assessment of the interannual variability and impact of the QBO and upwelling on tracer-tracer distributions of N_2O and O_3 in the tropical lower stratosphere, *Atmos. Chem. Phys.*, 13, 3619-3641, doi:10.5194/acp-13-3619-2013, 2013.

Kinnison, D. E., Brasseur, G. P., Walters, S., Garcia, R. R., Sassi, F., Boville, B. A., Marsh, D., Harvey, L., Randall, C., Randel, W., Lamarque, J.-F., Emmons, L. K., Hess, P., Orlando, J., Tyndall, J., and Pan, L.: Sensitivity of chemical tracers to meteorological parameters in the MOZART-3 chemical transport model, J. Geophys. Res., 112, D20302, doi:10.1029/2006JD007879, 2007.

Kunz, A., Pan, L., Konopka, P., Kinnison, D., and Tilmes, S.: Chemical and dynamical discontinuity at the extratropoical tropopause based on START08 and WACCM analysis, J. Geophys. Res., 116, D24302, doi:10.1029/2011JD016686, 2011.

Kuttippurath, J., Bodeker, G. E., Roscoe, H. K., and Nair, P. J.: A cautionary note on the use of EESC-based regression analysis for ozone trend studies, Geophys. Res. Lett., 42, 162-168, doi:10.1002/2014GL062142, 2015.

Kvissel, O.-K., Orsolini, Y. J., Stordal, F., Isaksen, I. S. A., and Santee, M. L.: Formation of stratospheric nitric acid by a hydrated ion cluster reaction: Implications for the effect of energetic particle precipitation on the middle atmosphere, J. Geophys. Res., 117, D16301, doi:10.1029/2011JD017257, 2012.

Lamarque, J.-F., Emmons, L. K., Hess, P. G., Kinnison, D. E., Tilmes, S., Vitt, F., Heald, C. L., Holland, E. A., Lauritzen, P. H., Neu, J., Orlando, J. J., Rasch, P. J., and Tyndall, G. K.: CAM-chem: description and evaluation of interactive atmospheric chemistry in the Community Earth System Model, Geosci. Model Dev., 5, 369–411, https://doi.org/10.5194/gmd-5-369-2012, 2012.

Lambert, A., Read, W. G., Livesey, N. J., Santee, M. L., Manney, G. L., Froidevaux, L., Wu, D. L., Schwartz, M. J., Pumphrey, H. C., Jimenez, C., Nedoluha, G. E., Cofield, R. E., Cuddy, D. T., Daffer, W. H., Drouin, B. J., Fuller, R. A., Jarnot, R. F., Knosp, B. W., Pickett, H. M., Perun, V. S., Snyder, W. V., Stek, P. C., Thurstans, R. P., Wagner, P. A., Waters, J. W., Jucks, K. W., Toon, G. C., Stachnik, R. A., Bernath, P. F., Boone, C. D., Walker, K. A., Urban, J., Murtagh, D., Elkins, J. W., and Atlas, E.: Validation of the Aura Microwave Limb Sounder stratospheric water vapour and nitrous oxide measurements, J. Geophys. Res., 112, D24S36, doi:10.1029/2007JD008724, 2007.

Lin, S.–J., A "vertically-Lagrangian" finite-volume dynamical core for global atmospheric models, Mon. Wea. Rev., 132, 2293-2307, 2004.

Livesey, N. J., Filipiak, M. J., Froidevaux, L., Read, W. G., Lambert, A., Santee, M. L., Jiang, J. H., Waters, J. W., Cofield,R. E., Cuddy, D. T., Daffer, W. H., Drouin, B. J., Fuller, R. A., Jarnot, R. F., Jiang, Y. B., Knosp, B. W., Li, Q. B., Perun, V. S., Schwartz, M. J., Snyder,W. V., Stek, P. C., Thurstans, R. P.,Wagner, P. A., Pumphrey, H. C., Avery, M., Browell, E. V., Cammas, J.-P., Christensen, L. E., Edwards, D. P., Emmons, L. K., Gao, R.-S., Jost, H.-J., Loewenstein, M., Lopez, J. D., Nedelec, P., Osterman, G. B., Sachse, G. W., and Webster, C. R.: Validation of Aura Microwave Limb Sounder O_3 and CO observations in the upper troposphere and lower stratosphere, J. Geophys. Res., 113, D15S02, doi:10.1029/2007JD008805, 2008.

Livesey, N. J., Read, W. G., Froidevaux, L., Lambert, A., Manney, G. L., Pumphrey, H. C., Santee, M. L., Schwartz, M. J., Wang, S., Cofield, R. E., Cuddy, D. T., Fuller, R. A., Jarnot, R. F., Jiang, J. H., Knosp, B. W., Stek, P. C., Wagner, P. A., and Wu, D. L.: EOS MLS Version 4.2x Level 2 data quality and description document, Tech. rep., Jet Propulsion Laboratory D-33509 Rev. D, available from http://mls.jpl.nasa.gov/, 2018.

Lopez-Puertas, M., Funke, B., Gil-Lopez, S., von Clarmann, T., Stiller, G. P., Höpfner, M., Kellmann, S., Mengistu Tsidu, G., Fischer, H., and Jackman, C. H.: HNO_3, N_2O_5 and $ClONO_2$ Enhancements after the October–November 2003 Solar Proton Events, J. Geophys. Res., 110, A09S44, doi:10.1029/2005JA011051, 2005b.

Lossow, S., Khosrawi, F., Nedoluha, G. E., Azam, F., Bramstedt, K., Burrows, J. P., Dinelli, B. M., Eriksson, P., Espy, P. J., García-Comas, M., Gille, J. C., Kiefer, M., Noël, S., Raspollini, P., Read, W. G., Rosenlof, K. H., Rozanov, A., Sioris, C. E., Stiller, G. P., Walker, K. A., and Weigel, K.: The SPARC water vapour assessment II: comparison of annual, semi-annual and quasi-biennial variations in stratospheric and lower mesospheric water vapour observed from satellites, Atmos. Meas. Tech., 10, 1111–1137, https://doi.org/10.5194/amt-10-1111-2017, 2017a.

Lossow, S., Garny, H., and Jöckel, P.: An "island" in the stratosphere – on the enhanced annual variation of water vapour in the middle and upper stratosphere in the southern tropics and subtropics, Atmos. Chem. Phys., 17, 11521–11539, https://doi.org/10.5194/acp-17-11521-2017, 2017b.

Mahieu, E., Chipperfield, M. P., Notholt, J., Reddmann, T., Anderson, J., Bernath, P. F., Blumenstock, T., Coffey, M. T., Dhomse, S. S., Feng, W., Franco, B., Froidevaux, L., Griffith, D. W. T., Hannigan, J. W., Hase, F., Hossaini, R., Jones, N. B., Morino, I., Murata, I., Nakajima, H., Palm, M., Paton-Walsh, C., Russell III, J. M., Schneider, M., Servais, C., Smale, D., and Walker, K. A.: Recent Northern Hemisphere stratospheric HCl increase due to atmospheric circulation changes, Nature, 515, 104-107, doi:10.1038/nature13857, 2014.

Marsh, D. R., Mills, M. J., Kinnison, D. E., Lamarque, J.-F., Calvo, N., and Polvani, L. M.: Climate change from 1850 to 2005 simulated in CESM1 (WACCM), Journal of Climate, 26 (19), doi:10.1175/JCLI-D-12-00558.1, 2013.

Matthes, K., Marsh, D. R., Garcia, R. R., Kinnison, D. E., Sassi, F., and Walters, S.: Role of the QBO in modulating the influence of the 11-year solar cycle on the atmosphere using constant forcings, Journal of Geophysical Research, 115, D18110. https://doi.org/10.1029/2009JD013020, 2010.

Millán, L. F., Livesey, N. J., Santee, M. L., Neu, J. L., Manney, G. L., and Fuller, R. A.: Case studies of the impact of orbital sampling on stratospheric trend detection and derivation of tropical vertical velocities: solar occultation vs. limb emission sounding, Atmos. Chem. and Phys., 16, 11 521–11 534, doi:10.5194/acp-16-11521-2016, 2016.

Morgenstern, O., Hegglin, M. I., Rozanov, E., O'Connor, F. M., Abraham, N. L., Akiyoshi, H., Archibald, A. T., Bekki, S., Butchart, N., Chipperfield, M. P., Deushi, M., Dhomse, S. S., Garcia, R. R., Hardiman, S. C., Horowitz, L. W., Joeckel, P., Josse, B., Kinnison, D., Lin, M., Mancini, E., Manyin, M. E., Marchand, M., Marecal, V., Michou, M., Oman, L. D., Pitari, G., Plummer, D. A., Revell, L. E., Saint-Martin, D., Schofield, R., Stenke, A., Stone, K., Sudo, K., Tanaka, T. Y., Tilmes, S., Yamashita, Y., Yoshida, K., and Zeng, G.: Review of the global models used within phase 1 of the Chemistry-Climate Model Initiative (CCMI), Geosci. Model Dev., 10, 639–671, https://doi.org/10.5194/gmd-10-639-2017, 2017.

Nair, P., Godin-Beekmann, S., Kuttippurath, J., Ancellet, G., Goutail, F., Pazmino, A., Froidevaux, L., Zawodny, J. M., Evans, R. D., Wang, H. J., Anderson, J., and Pastel, M.: Ozone trends derived from the total column and vertical profiles at a northern mid-latitude station, Atmos. Chem. Phys., 13, 10373-10384, www.atmos-chem-phys.net/13/10373/2013/, doi:10.5194/acp-13-10373-2013, 2013.

Nair, P. J., Froidevaux, L., Kuttippurath, J., Zawodny, J. M., Russell III, J. M., Steinbrecht, W., Claude, H., Leblanc, T., van Gijsel, J. A. E., Johnson, B., Swart, D. P. J., Thomas, A., Querel, R., Wang, R., and Anderson, J.: Subtropical and midlatitude ozone trends in the stratosphere: Implications for recovery, J. Geophys. Res., 120, 7247-7257, doi:10.1002/2014JD022371, 2015.

Nedoluha, G. E., Gomez, R. M., Hicks, B. C., Wrotny, J. E., Boone, C., and Lambert, A.: Water vapor measurements in the mesosphere from Mauna Loa over solar cycle 23, J. Geophys. Res., 114, D23303, doi:10.1029/2009JD012504, 2009.

Nisbet, E. G., Dlugokencky, E. J., Manning, M. R., Lowry, D., Fisher, R. E., France, J. L., Michel, S. E., Miller, J. B., White, J. W. C., Vaughn, B., Bousquet, P., Pyle, J. A., Warwick, N. J., Cain, M., Brownlow, R., Zazzeri, G., Lanoisellé, M., Manning, A. C., Gloor, E., Worthy, D. E. J., Brunke, E.-G., Labuschagne, C., Wolff, E. W., and Ganesan, A. L.: Rising atmospheric methane: 2007–2014 growth and isotopic shift, Global Biogeochem. Cycles, 30, 1356–1370, doi:10.1002/2016GB005406, 2016.

Oltmans, S. J., Vömel, H., Hofmann, D. J., Rosenlof, K. H., and Kley, D.: The increase in stratospheric water vapor from balloonborne, frostpoint hygrometer measurements at Washington, D.C., and Boulder, Colorado, Geophys. Res. Lett., 27, No. 21, 3453-3456, 2000.

Oram, D. E., et al.: A growing threat to the ozone layer from short-lived anthropogenic chlorocarbons, Atmos. Chem. Phys., 17, 11929–11941, https://doi.org/10.5194/acp-17-11929-2017, 2017.

Orsolini, Y. J., Manney, G. L., Santee, M. L., and Randall, C. E.: An upper stratospheric layer of enhanced HNO_3 following exceptional solar storms, Geophys. Res. Lett., 32, L12S01, doi:10.1029/2004GL021588, 2005.

Perliski, L. M., Solomon, S., and London, J.: On the interpretation of seasonal variations of stratospheric ozone, Planet. Space Sci., 37, 12, 1527-1538, 1989.

Popp, P. J., Marcy, T. P., Watts, L. A., Gao, R. S., Fahey, D. W., Weinstock, E. M., Smith, J. B., Herman, R. L., Troy, R. F., Webster, C. R., Christensen, L. E., Baumgardner, D. G., Voigt, C., Kärcher, B., Wilson, J. C., Mahoney, M. J., Jensen, E. J., and Bui, T. P.: Condensed-phase nitric acid in a tropical subvisible cirrus cloud, Geophys. Res. Lett., 34, L24812, doi:10.1029/2007GL031832, 2007.

Popp, P. J., et al.: Stratospheric correlation between nitric acid and ozone, J. Geophys. Res., 114, D03305, doi:10.1029/2008JD010875, 2009.

Randel, W. J., Wu, F., and Gaffen, D. J.: Interannual variability of the tropical tropopause derived from radiosonde data and NCEP reanalysis, J. Geophys. Res., 105, 15,509- 15,523, 2000.

Randel, W. J., Wu, F., Oltmans, S. J., Rosenlof, K., and Nedoluha, G. E.: Interannual changes of stratospheric water vapor and correlations with tropical tropopause temperatures, J. Atmos. Sci., 61, 2133–2148, 2004.

Randel, W. J., Wu, F., Vömel, H., Nedoluha, G. E., and Forster, P.: Decreases in stratospheric water vapor after 2001: Links to changes in the tropical tropopause and the Brewer–Dobson circulation, J. Geophys. Res., 111, D12312, doi:10.1029/2005JD006744, 2006.

Randel, W. J., and Thompson, A. M.: Interannual variability and trends in tropical ozone derived from SAGE II satellite data and SHADOZ ozonesondes, J. Geophys. Res., 116, D07303, doi:10.1029/2010JD015195, 2011.

Randel, W. J., and Jensen, E. J.: Physical processes in the tropical tropopause layer and their roles in a changing climate, Nat. Geosci.,6, 169–176, doi:10.1038/ngeo1733, 2013.

Ray, E. A., Holton, J. R., Fishbein, E., Froidevaux, L., and Waters, J.: The tropical semiannual oscillations in temperature and ozone as observed by the MLS, J. Atmos. Sci., 51, 3045-3052, doi:10.1175/1520-0469(1994)051<3045:TTSOIT>2.0.CO;2, 1994.

Read, W. G., Lambert, A., Bacmeister, J., Cofield, R. E., Christensen, L. E., Cuddy, D. T., Daffer, W. H., Drouin, B. J., Fetzer, E., Froidevaux, L., Fuller, R., Herman, R., Jarnot, R. F., Jiang, J. H., Jiang, Y. B., Kelly, K., Knosp, B. W., Kovalenko, L. J., Livesey, N. J., Liu, H.-C., Manney, G. L., Pickett, H. M., Pumphrey, H. C., Rosenlof, K. H., Sabounchi, X., Santee, M. L., Schwartz, M. J., Snyder, W. V., Stek, P. C., Su, H., Takacs, L. L., Thurstans, R. P., Vömel, H., Wagner, P. A., Waters, J. W., Web- ster, C. R., Weinstock, E. M., and Wu, D. L.: Aura Microwave Limb Sounder upper tropospheric and lower stratospheric H_2O and relative humidity with respect to ice validation, J. Geophys. Res., 112, D24S35, doi:10.1029/2007JD008752, 2007.

Read, W. G., Schwartz, M. J., Lambert, A., Su, H., Livesey, N. J., Daffer, W. H., and Booe, C. D.: The roles of convection, extratropical mixing, and in-situ freeze-drying in the Tropical Tropopause Layer, Atmos. Chem. Phys., 8, 6051–6067, doi:10.5194/acp-8-6051-2008, 2008.

Rienecker, M., Suarez, M. J., Gelaro, R., Todling, R., Bacmeister, J., Liu, E., Bosilovich, M. G., Schubert, S. D., Takacs, L., Kim, G.-K., Bloom, S., Chen, J., Collins, D., Conaty, A., da Silva, A., Gu, W., Joiner, J., Koster, R. D., Lucchesi, R., Molod, A., Owens, T., Pawson, S., Pegion, P., Redder, C. R., Reichle, R., Robertson, J., F. R., Ruddick, A. G., Sienkiewicz, M., and Woollen, J.: MERRA: NASA's Modern-Era Retrospective Analysis for Research and Applications, J. Climate, 24, 3624–3648, doi:10.1175/JCLI-D-11-00015.1, 2011.

Rohs, S., Schiller, C., Riese, M., Engel, A., Schmidt, U., Wetter, T., Levin, I., Nakazawa, T., and Aoki, S.: Long-term changes of methane and hydrogen in the stratosphere in the period 1978–2003 and their impact on the abundance of stratospheric water vapor, J. Geophys. Res., 111, D14315, doi:10.1029/2005JD006877, 2006.

Rosenlof, K. H., and Reid, G. R.: Trends in the temperature and water vapor content of the tropical lower stratosphere: Sea surface connection, J. Geophys. Res., 113, D06107, doi:10.1029/2007JD009109, 2008.

Ryan, N. J., Kinnison, D. E., Garcia, R. R., Hoffmann, C. G., Palm, M., Raffalski, U., and Notholt, J.: Assessing the ability to derive rates of polar middle-atmospheric descent using trace gas measurements from remote sensors, Atmos. Chem. Phys., 18, 1457–1474, https://doi.org/10.5194/acp-18-1457-2018, 2018.

Sakazaki, T., Shiotani, M., Suzuki, M., Kinnison, D., Zawodny, J. M., McHugh, M., and Walker, K. A.: Sunset-sunrise difference in solar occultation ozone measurements (SAGE II, HALOE, and ACE-FTS) and its relationship to tidal vertical winds, *Atmos. Chem. Phys.*, 15, 829-843, doi:10.5194/acp-15-829-2015, 2015.

Santee, M. L., Lambert, A., Read, W. G., Livesey, N. J., Cofield, R. E., Cuddy, D. T., Daffer, W. H., Drouin, B. J., Froidevaux, L., Fuller, R. A., Jarnot, R. F., Knosp, B. W., Manney, G. L., Perun, V. S., Snyder, W. V., Stek, P. C., Thurstans, R. P., Wagner, P. A., Waters, J. W., Muscari, G., de Zafra, R. L., Dibb, J. E., Fahey, D. W., Popp, P. J., Marcy, T. P., Jucks, K. W., Toon, G. C., Stachnik, R. A., Bernath, P. F., Boone, C. D., Walker, K.A., Urban, J., and Murtagh, D.: Validation of the Aura Microwave Limb Sounder HNO_3 measurements, J. Geophys. Res., 112, D24S40, doi:10.1029/2007JD008, 2007.

Santer, B. D., Wigley, T. M. L., Boyle, J. S., Gaffen, D. J., Hnilo, J. J., Nychka, D., Parker, D. E., and Taylor, K. E.: Statistical significance of trends and trend differences in layer-average atmospheric temperature time series, J. Geophys. Res., 105, 7337-7356, 2000.

Schaefer, H., Mikaloff Fletcher, S. E., Veidt, C., Lassey, K. R., Brailsford, G. W., Bromley, T. M., Dlugokencky, E. J., Michel, S. E., Miller, J. B., Levin, I., Lowe, D. C., Martin, R. J., Vaughn, B. H., and White, J. W. C.: A 21st century shift from fossil-fuel to biogenic methane emissions indicated by $^{13}CH_4$, Science, doi:10.1126/science.aad2705, 2016.

Scherer, M., Vömel, H., Fueglistaler, S., Oltmans, S. J., and Staehelin, J.: Trends and variability of midlatitude stratospheric water vapour deduced from the re-evaluated Boulder balloon series and HALOE, Atmos. Chem. Phys., 8, 1391-1402, https://doi.org/10.5194/acp-8-1391-2008, 2008.

Schoeberl, M. R., Douglass, A. R., Newman, P. A., Lait, L. R., Lary, D., Waters, J., Livesey, N., Froidevaux, L., Lambert, A., Read,W., Filipiak, M. J., and Pumphrey, H. C.: QBO and annual cycle variations in tropical lower stratosphere trace gases from HALOE and Aura MLS observations, J. Geophys. Res., 113, D05301, doi:10.1029/2007JD008678, 2008.

Schoeberl, M. R., Dessler, A. E., and Wang, T.: Modeling upper tropospheric and lower stratospheric water vapor anomalies, Atmos. Chem. Phys., 13, 7783–7793, https://doi.org/10.5194/acp-13-7783-2013, 2013.

Smith, A. K., Garcia, R. R., Moss, A. C., and Mitchell, N. J.: The semiannual oscillation of the tropical zonal wind in the middle atmosphere derived from satellite geopotential height retrievals, J. Atmos. Sci., 74, 2413-2425, doi:10.1175/JAS-D-17-0067.1, 2017.

Solomon, S., Kinnison, D. E., Bandoro, J., and Garcia, R.: Simulations of Polar Ozone Depletion: An Update, J. Geophys. Res., 120, 7958-7974, doi:10.1002/2015JD0233652015, 2015.

Solomon, S., Ivy, D. J., Kinnison, D., Mills, M. J., Neely III, R. R., and Schmidt, A.: Emergence of healing in the Antarctic ozone layer, Science, 353, 269–274, 2016.

SPARC, 2010: SPARC CCMVal Report on the Evaluation of Chemistry-Climate Models, Eyring, V., Shepherd, T., and Waugh, D. (Eds.), SPARC Report No. 5, WCRP-30/2010, WMO/TD – No. 40, available at www.sparc-climate.org/publications/sparc-reports/, 2010.

SPARC, 2017: The SPARC Data Initiative: Assessment of stratospheric trace gas and aerosol climatologies from satellite limb sounders, M. I. Hegglin and S. Tegtmeier (Eds.), SPARC Report No. 8, WCRP-05/2017, doi:10.3929/ethz-a-010863911, available at www.sparc-climate.org/publications/sparc-reports/, 2017.

Steinbrecht, W., Froidevaux, L., Fuller, R., Wang, R., Anderson, J., Roth, C., Bourassa, A., Degenstein, D., Damadeo, R., Zawodny, J., Frith, S., McPeters, R., Bhartia, P., Wild, J., Long, C., Davis, S., Rosenlof, K., Sofieva, V., Walker, K., Rahpoe, N., Rozanov, A., Weber, M., Laeng, A., von Clarmann, T., Stiller, G., Kramarova, N., Godin-Beekmann, S., Leblanc, T., Querel, R., Swart, D., Boyd, I., Hocke, K., Kämpfer, N., Maillard Barras, E., Moreira, L., Nedoluha, G., Vigouroux, C., Blumenstock, T., Schneider, M., Garcia, O., Jones, N., Mahieu, E., Smale, D., Kotkamp, M., Robinson, J., Petropavlovskikh, I., Harris, N., Hassler, B., Hubert, D. and Tummon, F.: An update on ozone profile trends for the period 2000 to 2016, Atmos. Chem. Phys. Discuss., 2017, 1–24, doi:10.5194/acp-2017-391, 2017.

Stiller, G. P., Mengistu Tsidu, G., von Clarmann, T., Glatthor, N., Höpfner, M., Kellmann, S., Linden, A., Ruhnke, R., Fischer, H., Lopez-Puertas, M., Funke, B., and Gil-Lopez, S.: An enhanced HNO_3 second maximum in the Antarctic midwinter upper stratosphere 2003, J. Geophys. Res., 110, D20303, doi:10.1029/2005JD006011, 2005.

Stolarski, R., Douglass, A., and Strahan, S. E.: Using satellite measurements of N_2O to remove dynamical variability from HCl measurements, Atmos. Chem. Phys., 18, 5691–5697, https://doi.org/10.5194/acp-18-5691-2018, 2018.

Stone, K. A., Solomon, S., and Kinnison, D. E.: On the identification of ozone recovery, Geophys. Res. Lett., 45, 5158-5165, doi:10.1029/2018GL077955, 2018.

Tegtmeier, S., Hegglin, M. I., Anderson, J., Bourassa, A., Brohede, S., Degenstein, D., Froidevaux, L., Fuller, R., Funke, B., Gille, J., Jones, A., Kasai, Y., Krüger, K., Kyrölä, E., Lingenfelser, G., Lumpe, J., Nardi, B., Neu, J., Pendlebury, D., Remsberg, E., Rozanov, A., Smith, L., Toohey, M., Urban, J., von Clarmann, T., Walker, K. A., and Wang, H. J.: The SPARC Data Initiative: A comparison of ozone climatologies from international satellite limb sounders, J. Geophys. Res. Atmos., 118, 12,229–12,247, doi: 10.1002/2013JD019877, 2013.

Tian, W., and Chipperfield, M. P.: Stratospheric water vapor trends in a coupled chemistry-climate model, Geophys. Res. Lett., 33, L06819, doi:10.1029/2005GL024675, 2006.

Tiao, G. C., Reinsel, G. C., Xu, D., Pedrick, J. H., Zhu, X., Miller, A. J., DeLuisi, J. J., Mateer, C. L., and Wuebbles, D. J.: Effects of autocorrelation and temporal sampling schemes on estimates of trend and spatial correlation, J. Geophys Res., 95, 20,507-20,517, 1990.

Tilmes, S., Lamarque, J.-F., Emmons, L. K., Kinnison, D., Marsh, D., Garcia, R. R., Smith, A. K., Neeley, R. R., Conley, A., Vitt, F., Val Martin, M., Tanimoto, H., Simpson, I., Blake, D. R., and Blake, N.: Representation of the Community Earth System Model (CESM1) CAM4-Chem within the Chemistry-Climate Model Initiative, Geosci. Model Dev., 9, 1853-1890, doi:10.5194/gmd-9-1853-2016, 2016.

Toohey, M., Hegglin, M. I., Tegtmeier, S., Anderson, J., Añel, J. A., Bourassa, A., Brohede, S., Degenstein, D., Froidevaux, L., Fuller, R., Funke, B., Gille, J., Jones, A., Kasai, Y., Krüger, K., Kyrölä, E., Neu, J. L., Rozanov, A., Smith, L., Urban, J., von Clarmann, T., Walker, K. A., and Wang, R.: Characterizing sampling bias in the trace gas climatologies of the SPARC Data Initiative, J. Geophys. Res. Atmos., 118, 11,847–11,862, doi: 10.1002/jgrd.5087, 2013.

Tummon, F., Hassler, B., Harris, N. R. P., Staehelin, J., Steinbrecht, W., Anderson, J., Bodeker, G. E., Bourassa, A., Davis, S. M., Degenstein, D., Frith, S. M., Froidevaux, L., Kyrölä, E., Laine, M., Long, C., Penckwitt, A. A., Sioris, C. E., Rosenlof, K. H., Roth, C., Wang, H.-J., and Wild, J.: Intercomparison of vertically resolved merged satellite ozone data sets: interannual variability and long-term trends, Atmos. Chem. Phys., 15, 3021-3043, https://doi.org/10.5194/acp-15-3021-2015, 2015.

Turner, A. J., Frankenberg, C., Wennberg, P. O., and Jacob, D. J.: Ambiguity in the causes for decadal trends in atmospheric methane and hydroxyl, Proc. Natl Acad. Sci. USA, 114, 5367-5372, www.pnas.org/cgi/doi/10.1073/pnas.1616020114, 2017.

Urban, J., Lossow, S., Stiller, G., and Read, W.: Another drop in water vapor, EOS Transactions, American Geophysical Union, 95, 27, 245-252, doi:10.1002/2014EO270001, 2014.

Verronen, P. T., Santee, M. L., Manney, G. L., Lehmann, R., Salmi, S.-M. and A. Seppälä.: Nitric acid enhancements in the mesosphere during the January 2005 and December 2006 solar proton events, J. Geophys. Res., 116, D17301, doi:10.1029/2011JD016075, 2011.

von Clarmann, T., Glatthor, N., H¨opfner, M., Kellmann, S., Ruhnke, R., Stiller, G. P., and Fischer, H.: Experimental evidence of perturbed odd hydrogen and chlorine chemistry after the October 2003 solar proton events, J. Geophys. Res., 110, A09S45, doi:10.1029/2005JA011053, 2005.

Vömel, H., Barnes, J. E., Forno, R. N., Fujiwara, M., Hasebe, F., Iwasaki, S., Kivi, R., Komala, N., Kyrölä, E., Leblanc, T., Morel, B., Ogino, S.-Y., Read, W. G., Ryan, S. C., Saraspriya, S., Selkirk, H., Shiotani, M., Valverde Canossa, J., and Whiteman, D. N.: Validation of Aura Microwave Limb Sounder water vapor by balloon-borne Cryogenic Frost point Hygrometer measurements, J. Geophys. Res., 112, D24S37, doi:10.1029/2007JD008698, 2007.

Waters, J. W., Froidevaux, L., Read, W. G., Manney, G. L., Eslon, L. S., Flower, D. A., Jarnot, R. F., and Harwood, R. S.: Stratospheric ClO and ozone from the Microwave Limb Sounder on the Upper Atmosphere Research Satellite, Nature, 362, 597-602, 1993.

Waters, J. W., Froidevaux, L., Harwood, R. S., Jarnot, R. F., Pickett, H. M., Read, W. G., Siegel, P. H., Cofield, R. E., Filipiak, M. J., Flower, D. A., Holden, J. R., Lau, G. K., Livesey, N. J., Manney, G. L., Pumphrey, H. C., Santee, M. L., Wu, D. L., Cuddy, D. T., Lay, R. R., Loo, M. S., Perun, V. S., Schwartz, M. J., Stek, P. C., Thurstans, R. P., Boyles, M. A., Chandra, S., Chavez, M. C., Chen, G.-S., Chudasama, B. V., Dodge, R., Fuller, R. A., Girard, M. A., Jiang, J. H., Jiang, Y., Knosp, B. W., LaBelle, R. C., Lam, J. C., Lee, K. A., Miller, M., Oswald, J. E., Patel, N. C., Pukala, D. M., Quintero, O., Scaff, D. M., Snyder, W. V., Tope, M. C., Wagner, P. A., and Walch, M. J.: The Earth Observing System Microwave Limb Sounder (EOS MLS) on the Aura satellite, IEEE Trans. Geosci. Remote Sens., 44 (5), 1075–1092, doi:10.1109/TGRS.2006.873771, 2006.

Waugh, D. W., Considine, D. B., and Fleming, E. L.: Is upper stratospheric chlorine decreasing as expected?, Geophys. Res. Lett., 28(7), 1187–1190, doi:10.1029/2000GL011745, 2001.

Waugh, D. W., and Eyring, V.: Quantitative performance metrics for stratospheric-resolving chemistry-climate models, Atmos. Chem. Phys., 8, 5699–5713, www.atmos-chem-phys.net/8/5699/2008/, 2008.

Weatherhead, E. C., Reinsel, G. C., Tiao, G. C., Meng, X.-L., Choi, D., Cheang, W.-K., Keller, T., DeLuisi, J., Wuebbles, D. J., Kerr, J. B., Miller, A. J., Oltmans, S. J., and Frederick, J. E.: Factors affecting the detection of trends: Statistical considerations and applications to environmental data, J. Geophys. Res.-Atmos., 103, 17149–17161, https://doi.org/10.1029/98JD00995, 1998.

Wegner, T., Kinnison, D. E., Garcia, R. R., and Solomon, S.: Simulation of polar stratospheric clouds in the specified dynamics version of the whole atmosphere community climate model, J. Geophys. Res., 4991–5002, https://doi.org/10.1002/jgrd.50415, 2013.

Wilka, C., Shah, K., Stone, K., Solomon, S., Kinnison, D., Mills, M., Schmidt, A., and Neely, R. R., III: The role of heterogeneous chemistry in ozone depletion and recovery, *Geophys. Res. Lett.*, 45, 7835-7842, doi:10.1029/2018GL078596, 2018.

WMO: Scientific Assessment of Ozone Depletion: 2014, Global Ozone Research and Monitoring Project-Report No. 55, WMO (World Meteorological Organization), Geneva, Switzerland, available at: https://www.esrl.noaa.gov/csd/assessments/ozone, 2014.

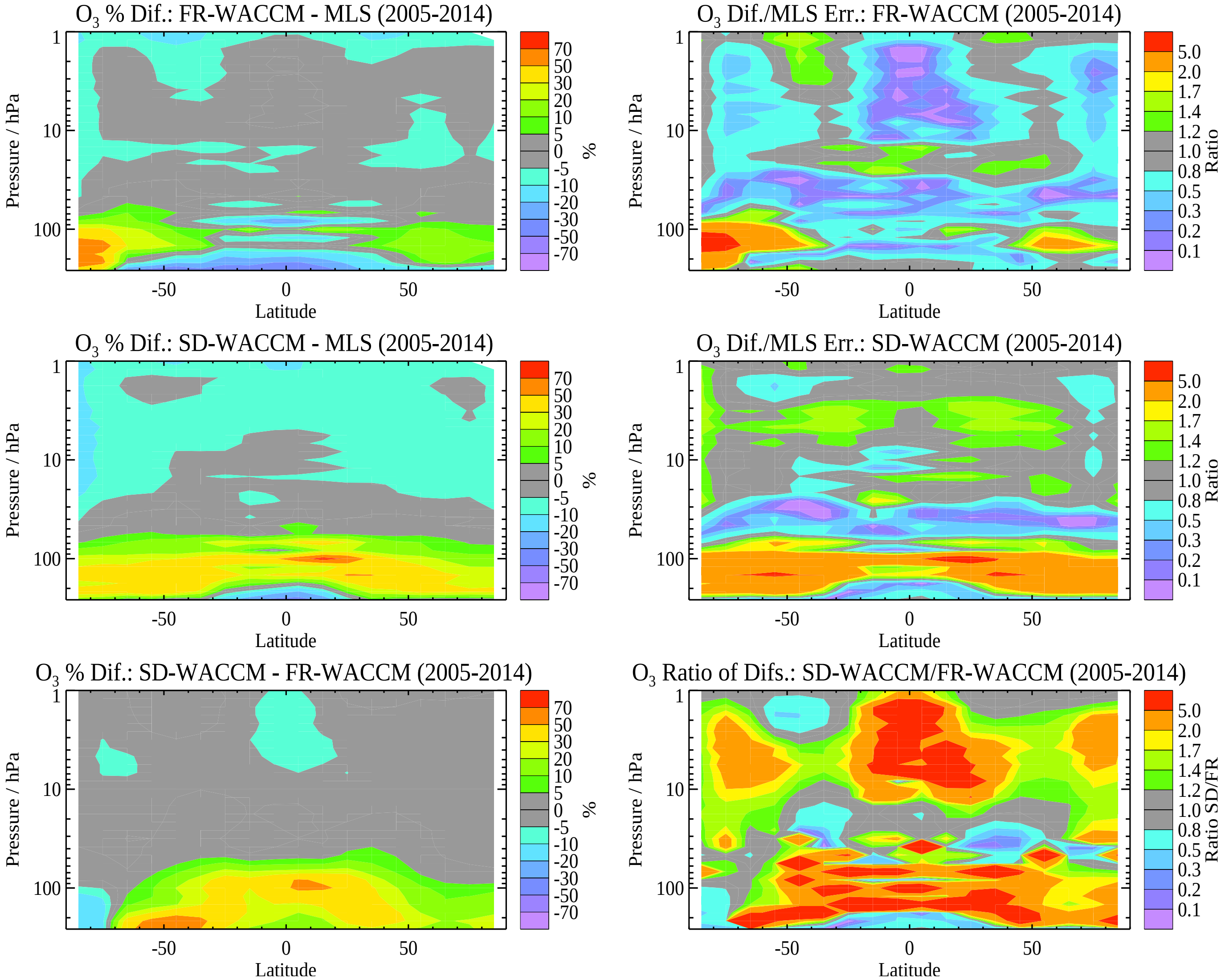

Figure 1. The left panels show percent differences ((model-data) divided by data) for binned climatological average O_3 from 2005 through 2014 (see Fig. S1 for these averages), for (top panel) the free-running model (average of 3 realizations) FR-WACCM, (middle panel) the specified dynamics model version SD-WACCM, and (bottom panel) SD-WACCM minus FR-WACCM, also as a percent difference. The two top right panels give ratios of the absolute value of average model (FR-WACCM in top panel, SD-WACCM in middle panel) minus average MLS O_3 to the MLS systematic O_3 errors, based on the climatological fields and estimated MLS errors (2σ estimates, see text and Fig. S2). The bottom right panel gives the ratios of average absolute differences between SD-WACCM and MLS to those differences for FR-WACCM and MLS; e.g., this shows that upper stratospheric tropical SD-WACCM mean O_3 is larger than O_3 from FR-WACCM, which is why SD-WACCM matches MLS better there (see top 2 right panels).

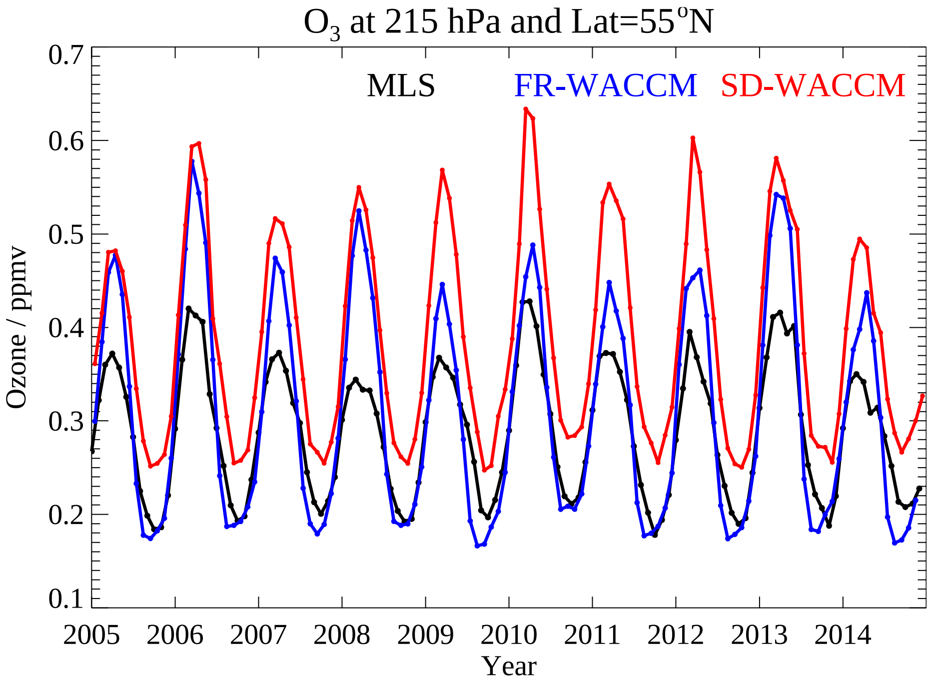

Figure 2. Monthly mean ozone time series at 215 hPa and the 55°N latitude bin (for averages over 50°N-60°N) from MLS, FR-WACCM, and SD-WACCM (see legend for color coding).

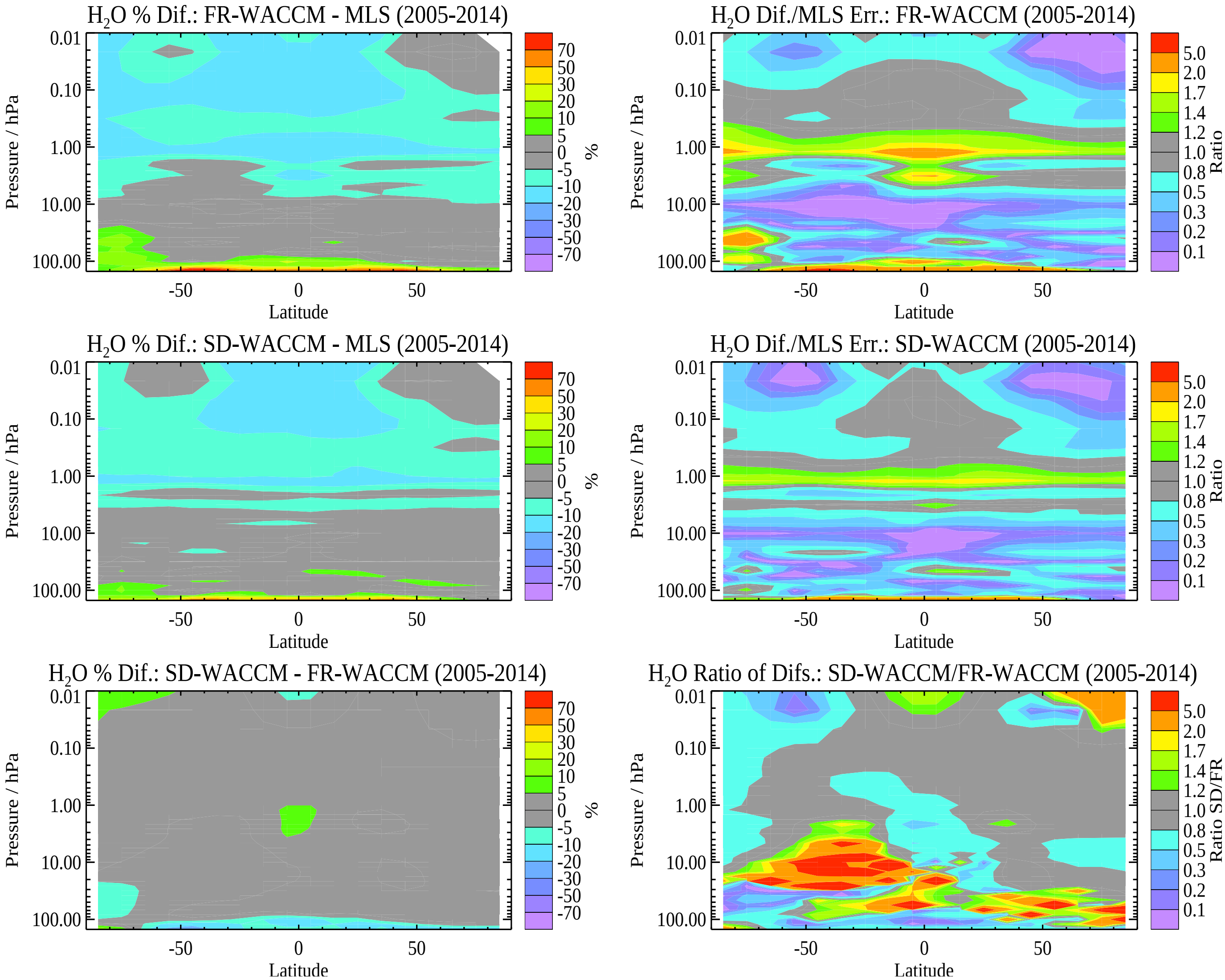

Figure 3. Same as Fig.1, but for stratospheric and mesospheric water vapor.

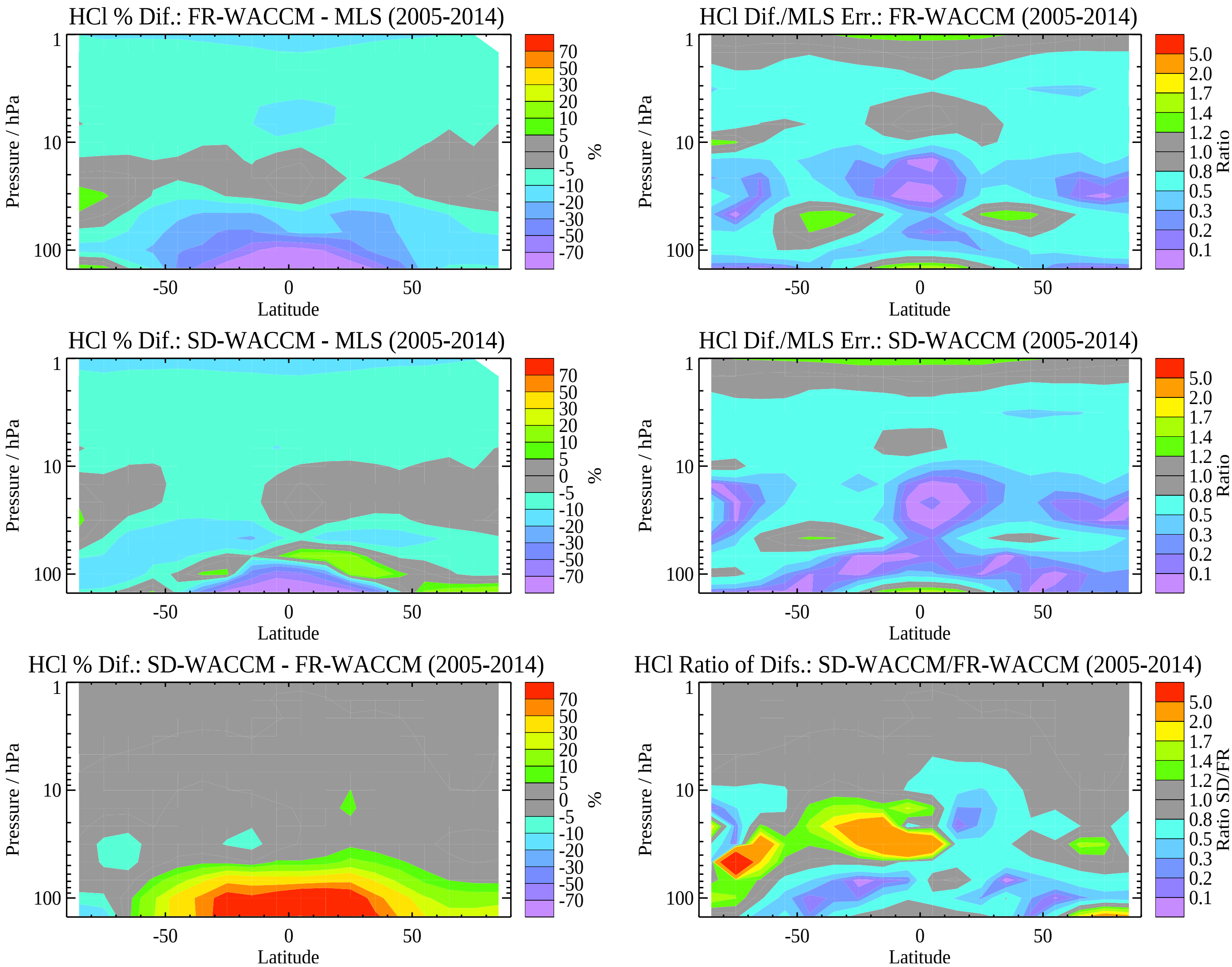

Figure 4. Same as Fig. 1, but for stratospheric HCl.

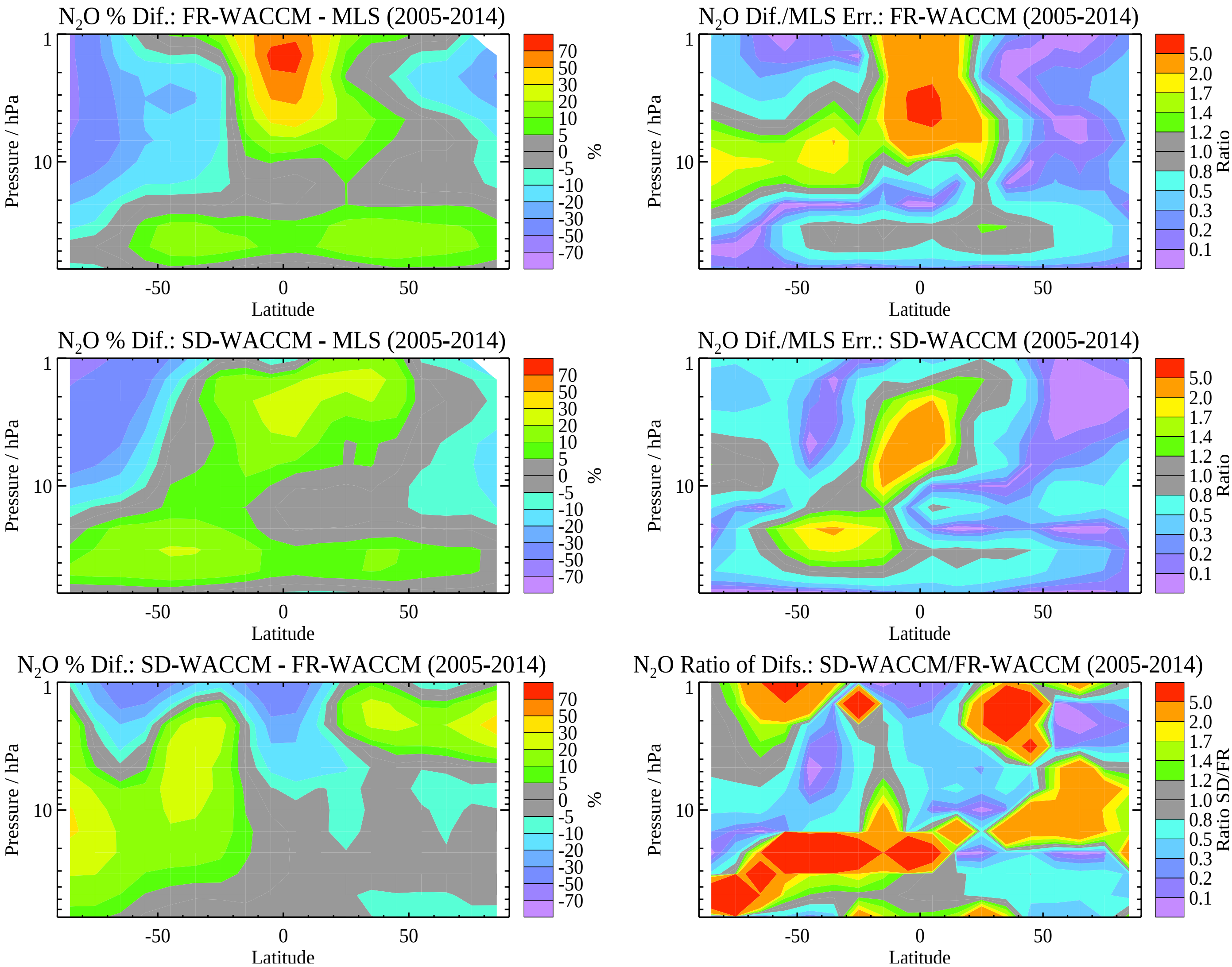

Figure 5. Same as Fig. 1, but for stratospheric N_2O.

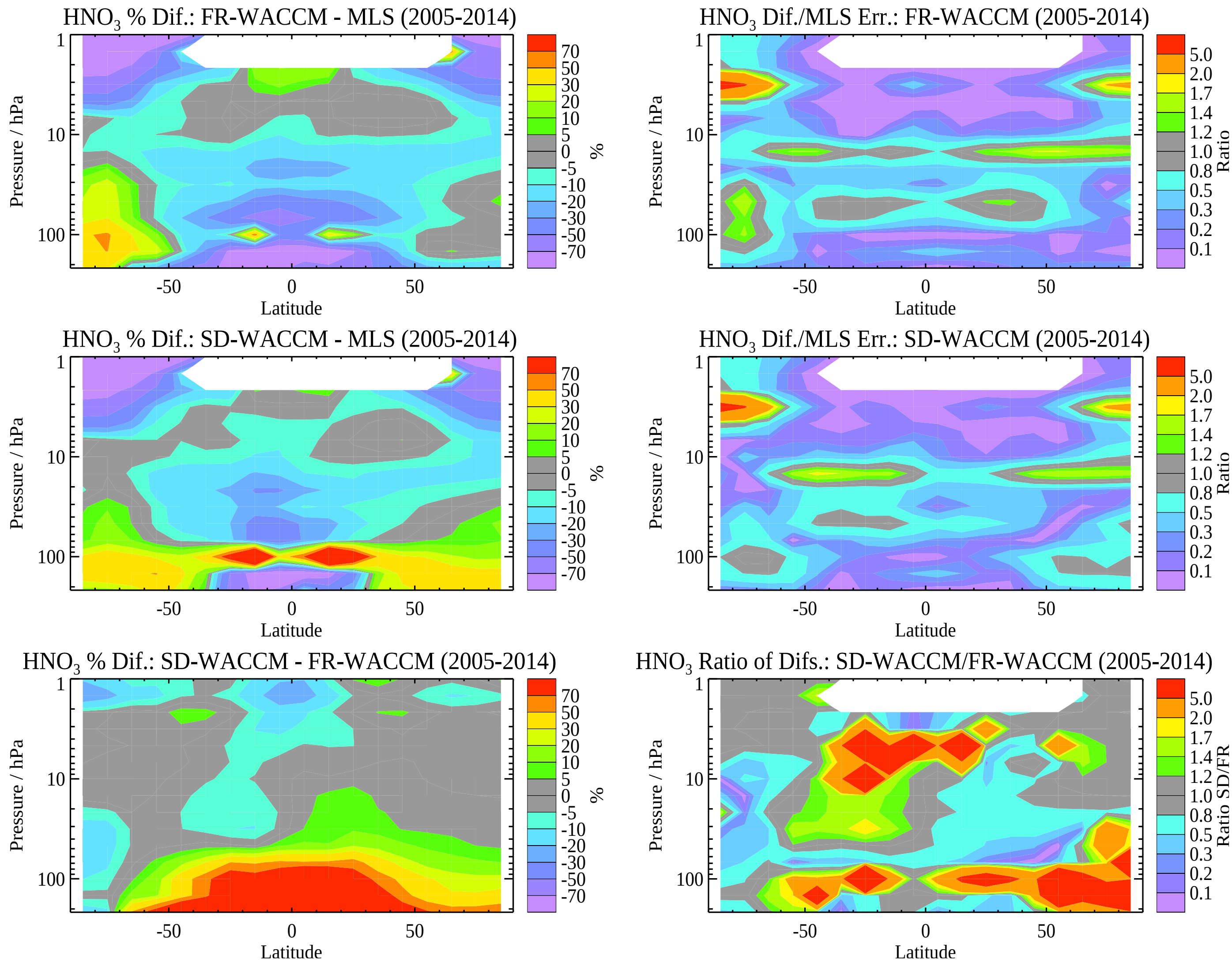

Figure 6. Same as Fig. 1, but for stratospheric HNO_3.

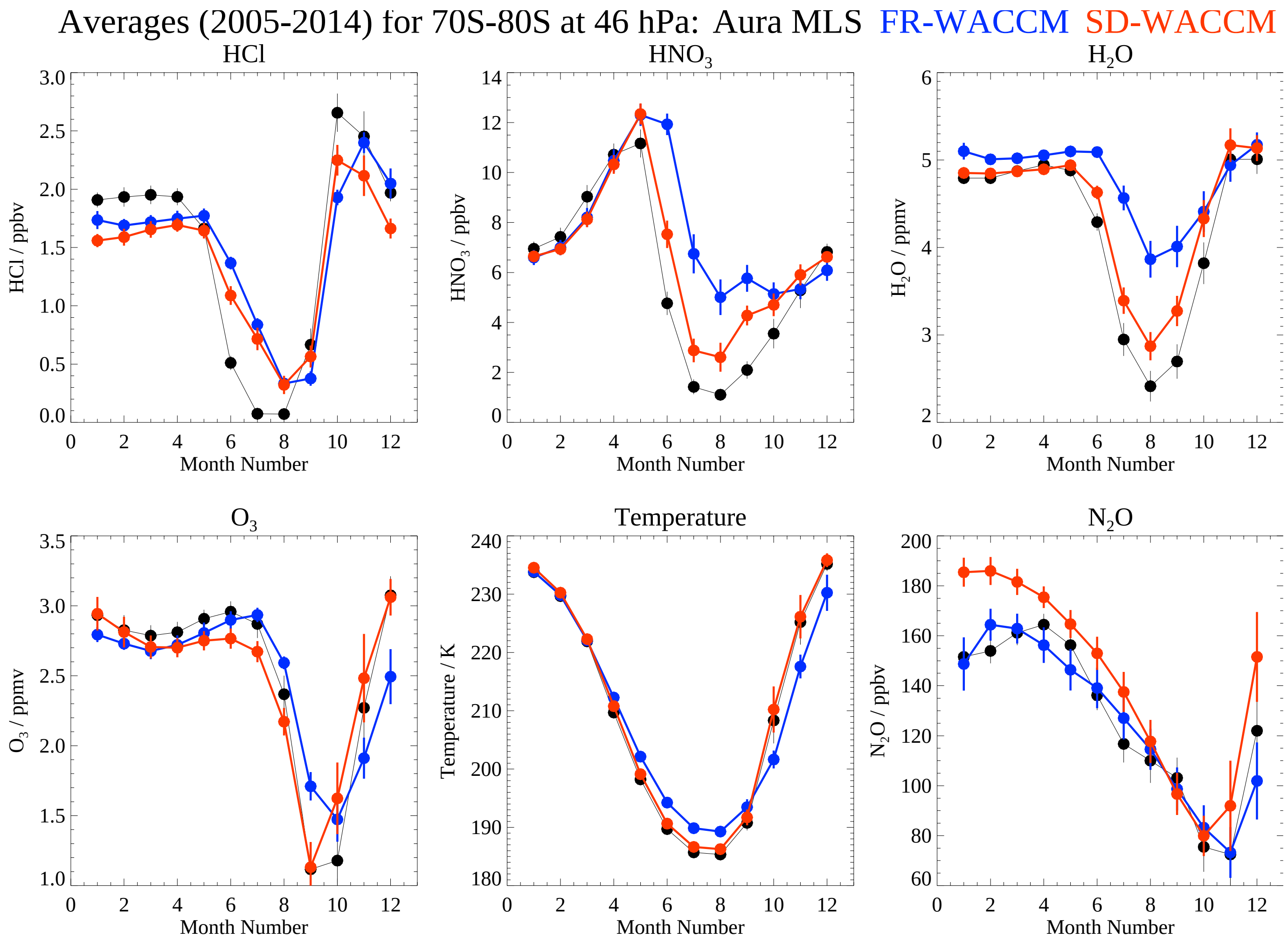

Figure 7. Each of the panels shows average seasonal changes from 2005-2014 for the 70°S-80°S region at 46 hPa. Data values (black) are from Aura MLS and model comparisons (FR-WACCM in blue, SD-WACCM in red) are provided for HCl (top left), HNO_3 (top center), H_2O (top right), O_3 (bottom left), temperature (bottom center), and N_2O (bottom right). For each month, the error bars represent twice the standard errors in the means, based on the set of 10 monthly averages (from 2005 through 2014).

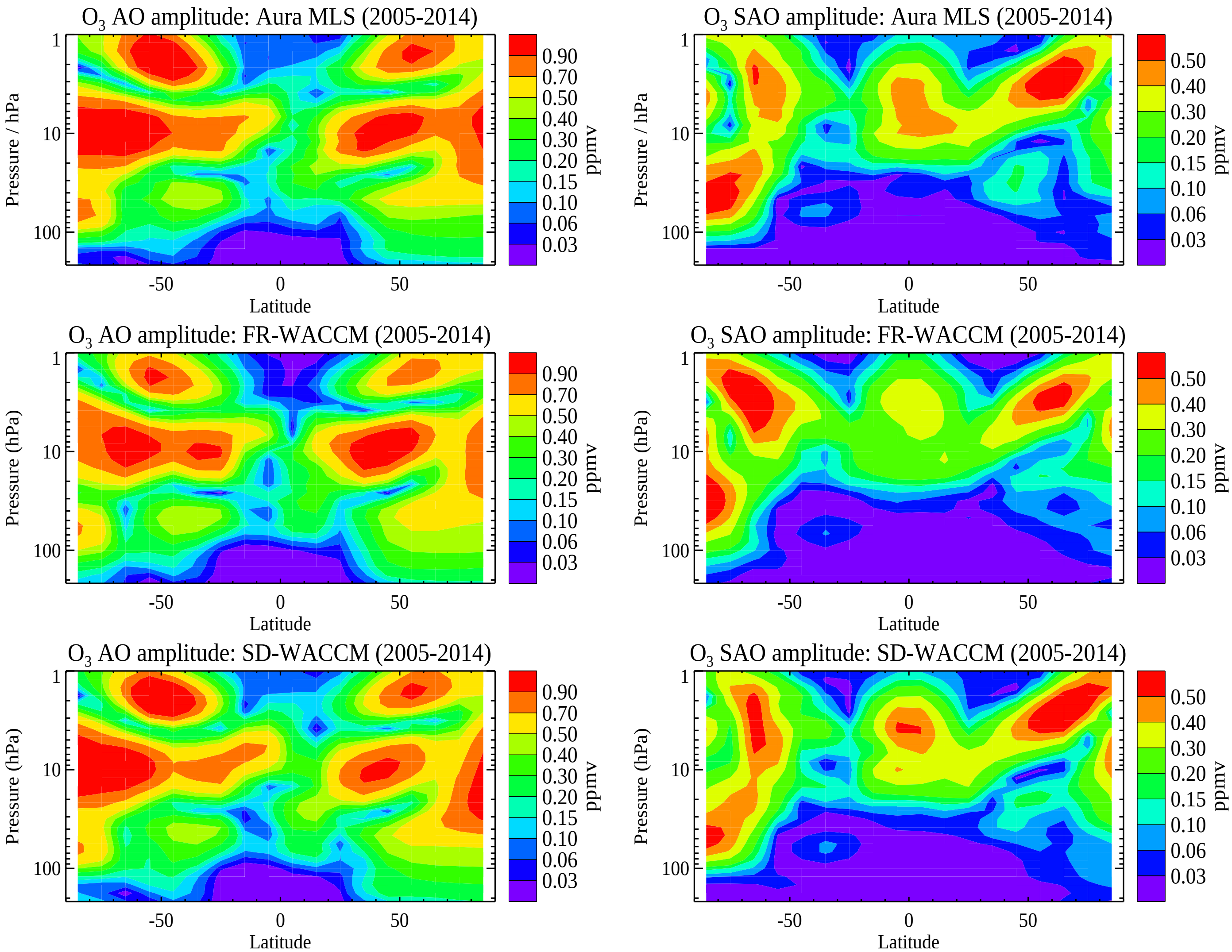

Figure 8. Amplitude of the stratospheric ozone annual cycle (left panels) and semi-annual cycle (right panels) for Aura MLS (top), FR-WACCM (middle), and SD-WACCM (bottom), based on fits to time series from 2005 through 2014.

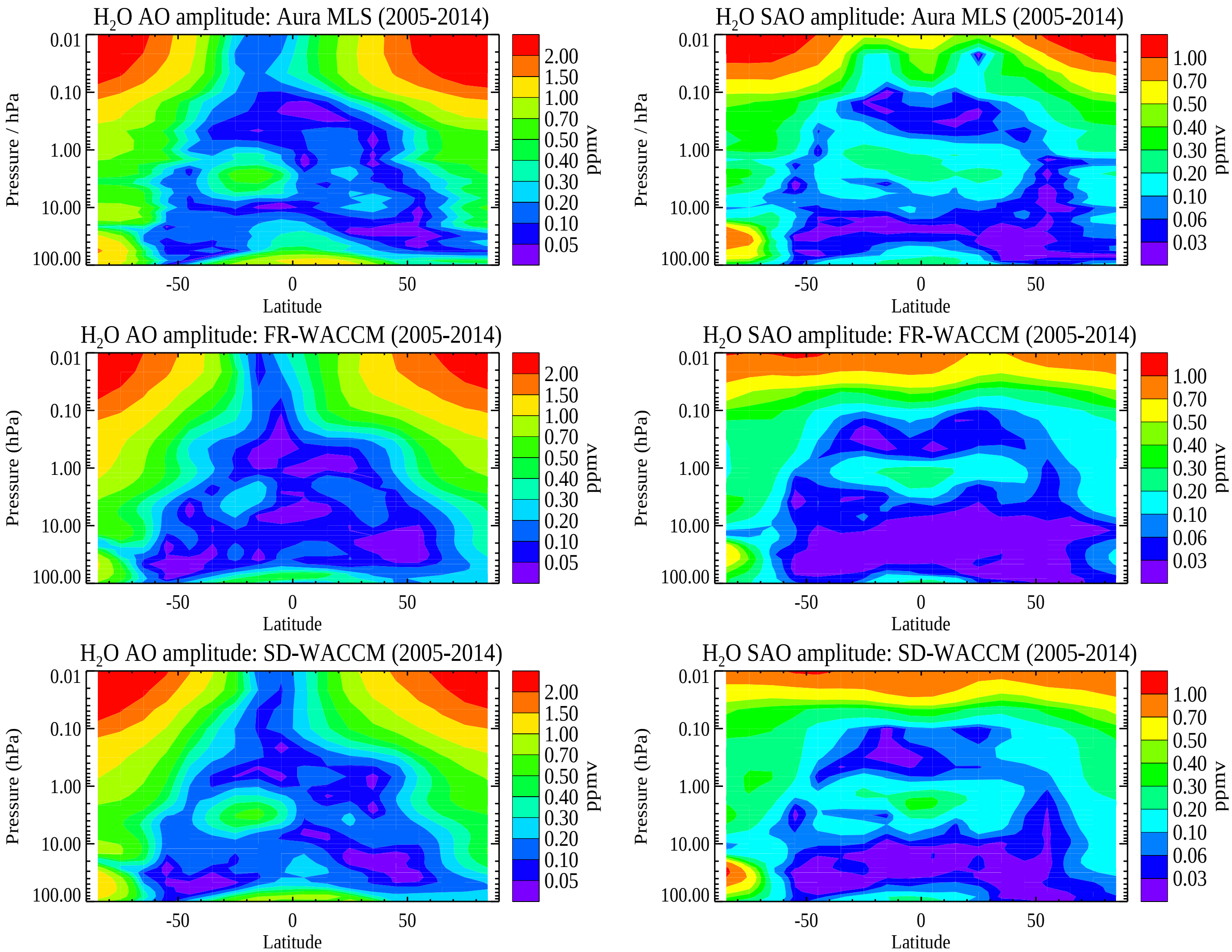

Figure 9. Same as Fig. 8, but for H_2O annual and semi-annual cycles in the stratosphere and mesosphere.

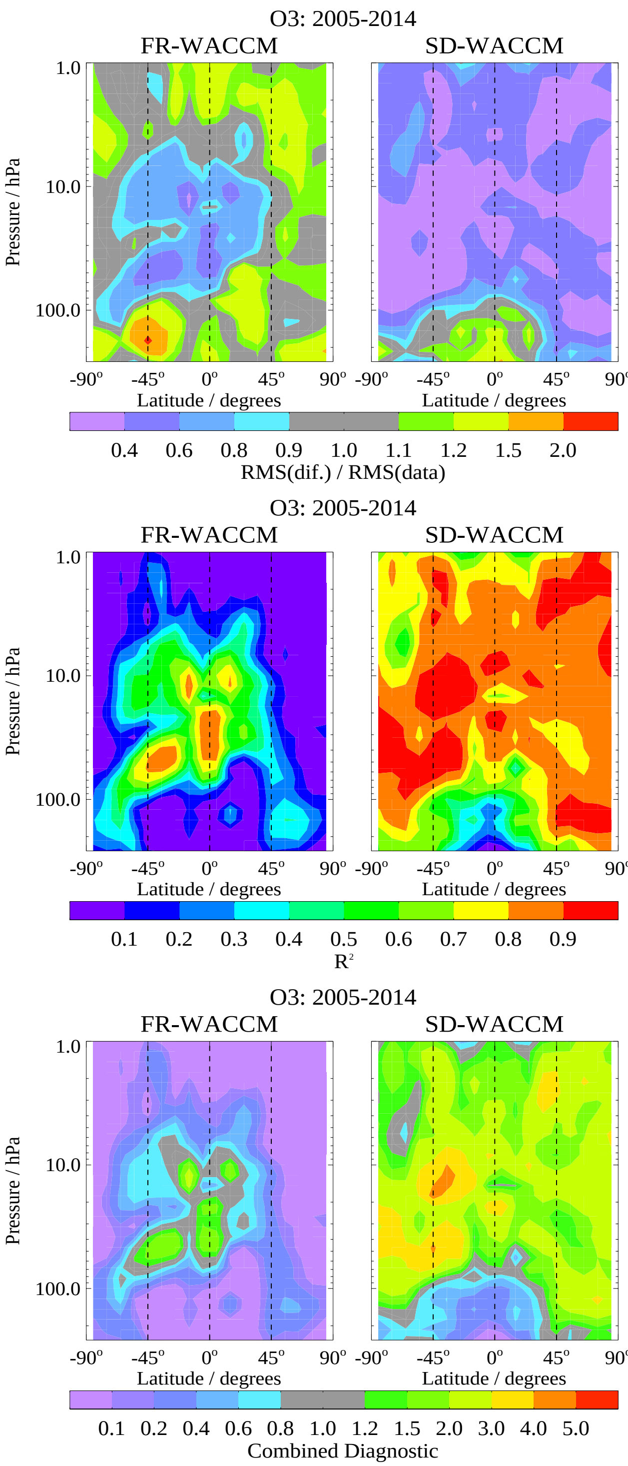

Figure 10. Latitude/pressure contours of diagnostics that show how well the deseasonalized anomalies of model ozone time series (FR-WACCM at left, SD-WACCM, at right) compare to MLS O_3 anomaly series for 2005-2014. Top panels show the RMS difference diagnostic (see text) and middle panels show R^2 values; small RMS difference values represent a closer fit, while large R^2 values represent highly correlated results. The bottom panels provide a combined diagnostic, namely the ratio of R^2 to the RMS difference diagnostic from the top panels; larger values here represent a better result for comparisons to the observed time series.

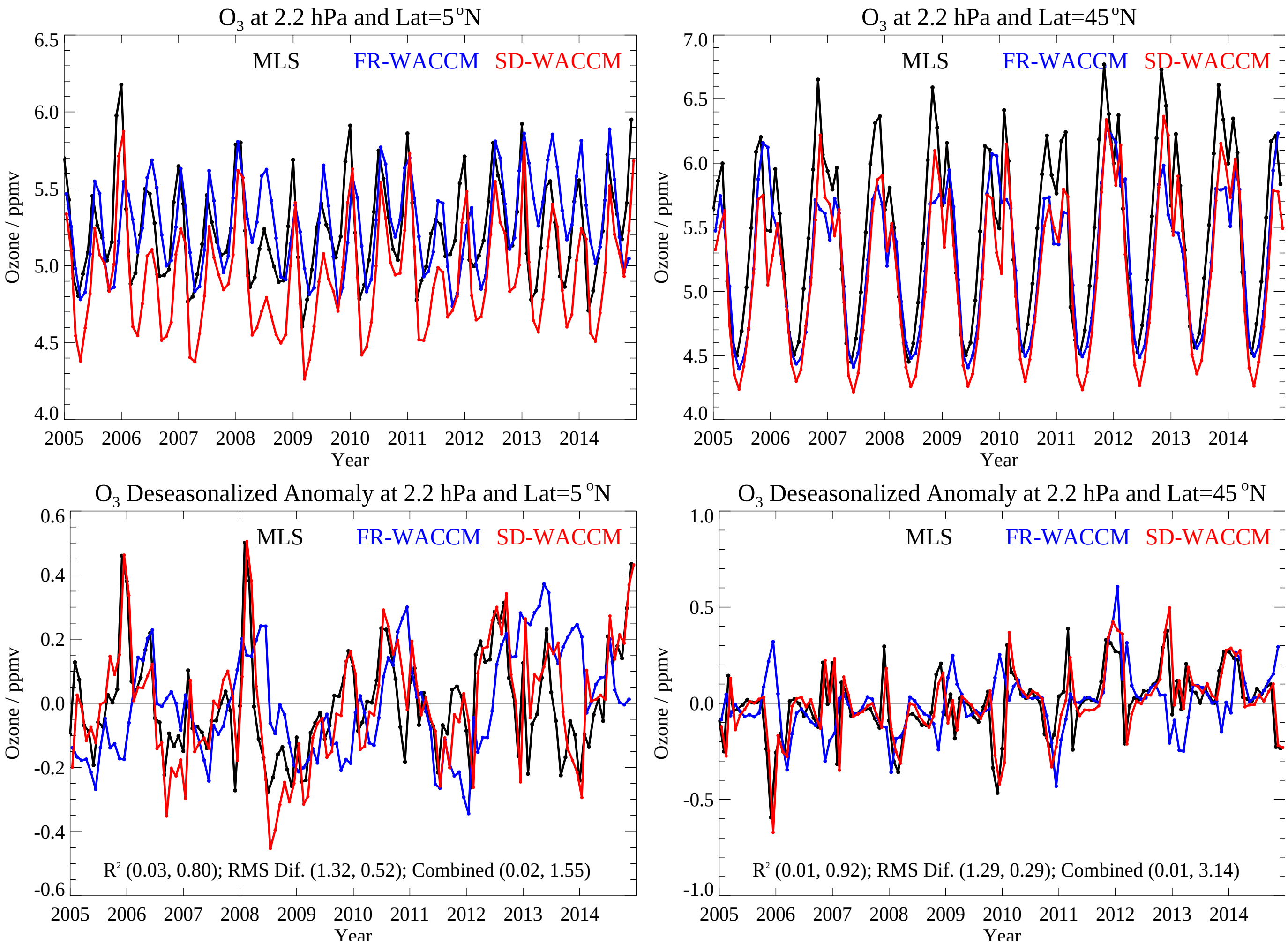

Figure 11. Time series of monthly zonal mean O_3 mixing ratios at 2.2 hPa (top panels) and deseasonalized anomalies (bottom panels), with the 0-10°N and 40°N-50°N latitude bins on the left and right, respectively. The two model time series (FR-WACCM in blue and SD-WACCM in red) are compared to the MLS series (in black) for 2005-2014. Diagnostic values (see text for a description) are shown in parentheses in the bottom two panels, with the 1st number referring to FR-WACCM and the 2nd number to SD-WACCM.

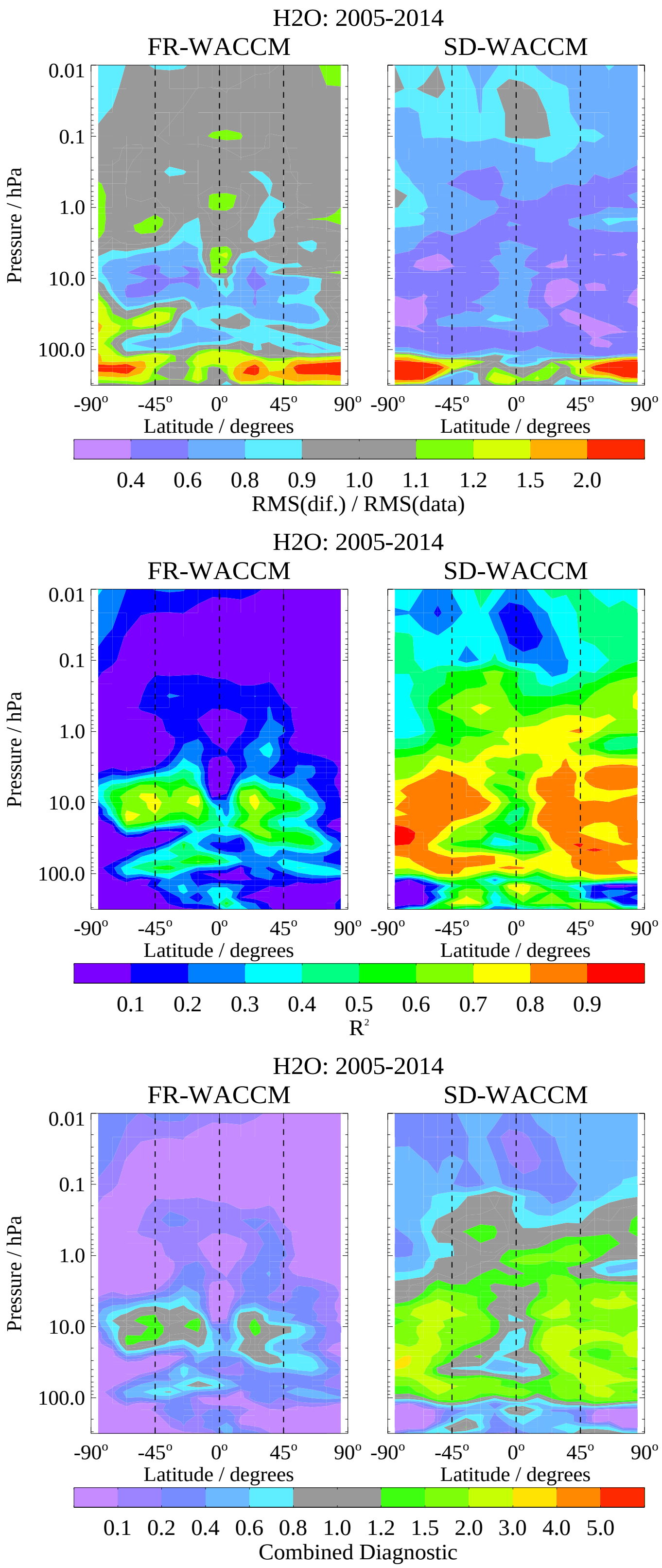

Figure 12. Same as the Fig. 10 diagnostics, but for H_2O up to 0.01 hPa.

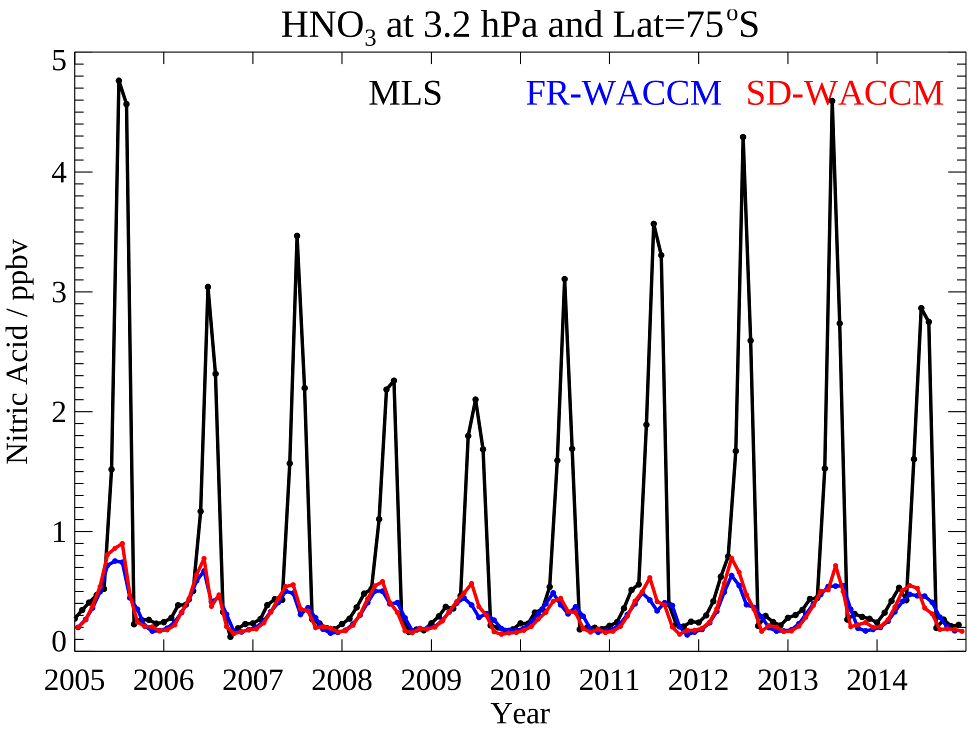

Figure 13. HNO_3 monthly zonal mean mixing ratio time series (2005 through 2014) from MLS, FR-WACCM, and SD-WACCM for 3.2 hPa and 70°S-80°S.

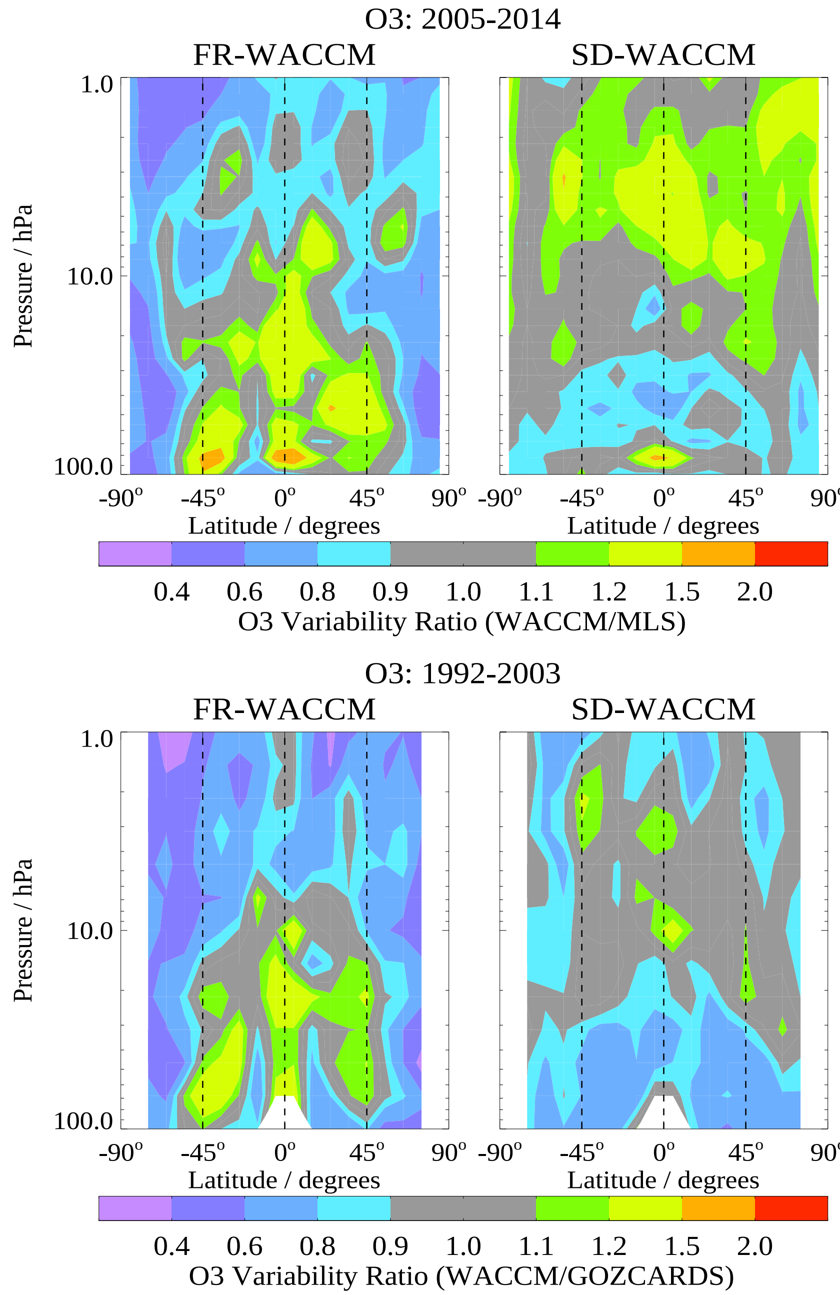

Figure 14. Variability ratios (model results divided by data results) for stratospheric O_3, with FR-WACCM results on the left, and SD-WACCM on the right. Before calculating the ratios, the variability values are obtained as the root mean square of detrended deseasonalized monthly anomaly time series, and expressed as a percentage of mean (climatological) abundances; the top panels show comparisons to MLS data for 2005-2014, whereas the bottom panels are for 1992-2003 comparisons to GOZCARDS.

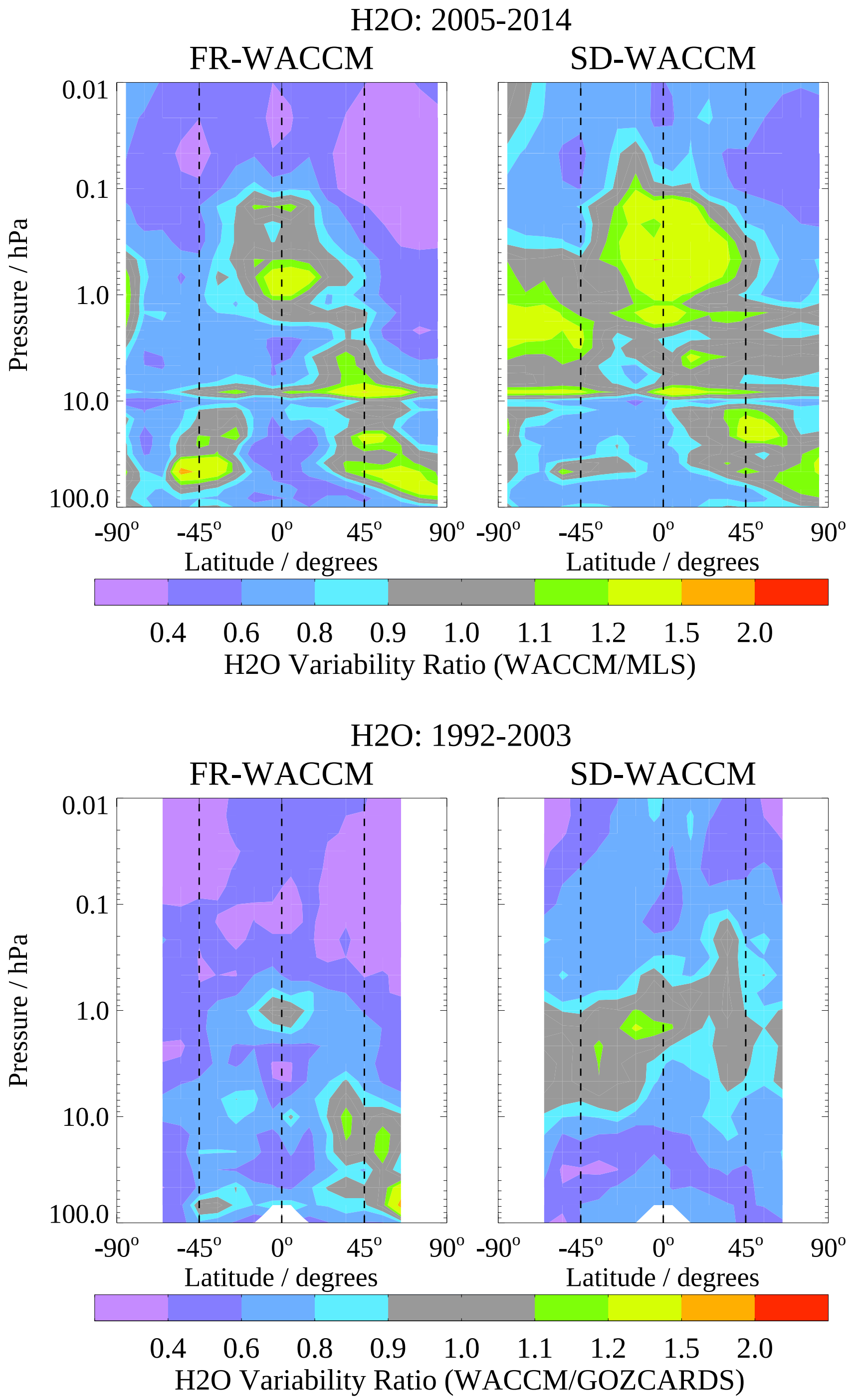

Figure 15. Same as Fig. 14, but for ratios (model/data) of H_2O stratospheric and mesospheric variability for two different time periods.

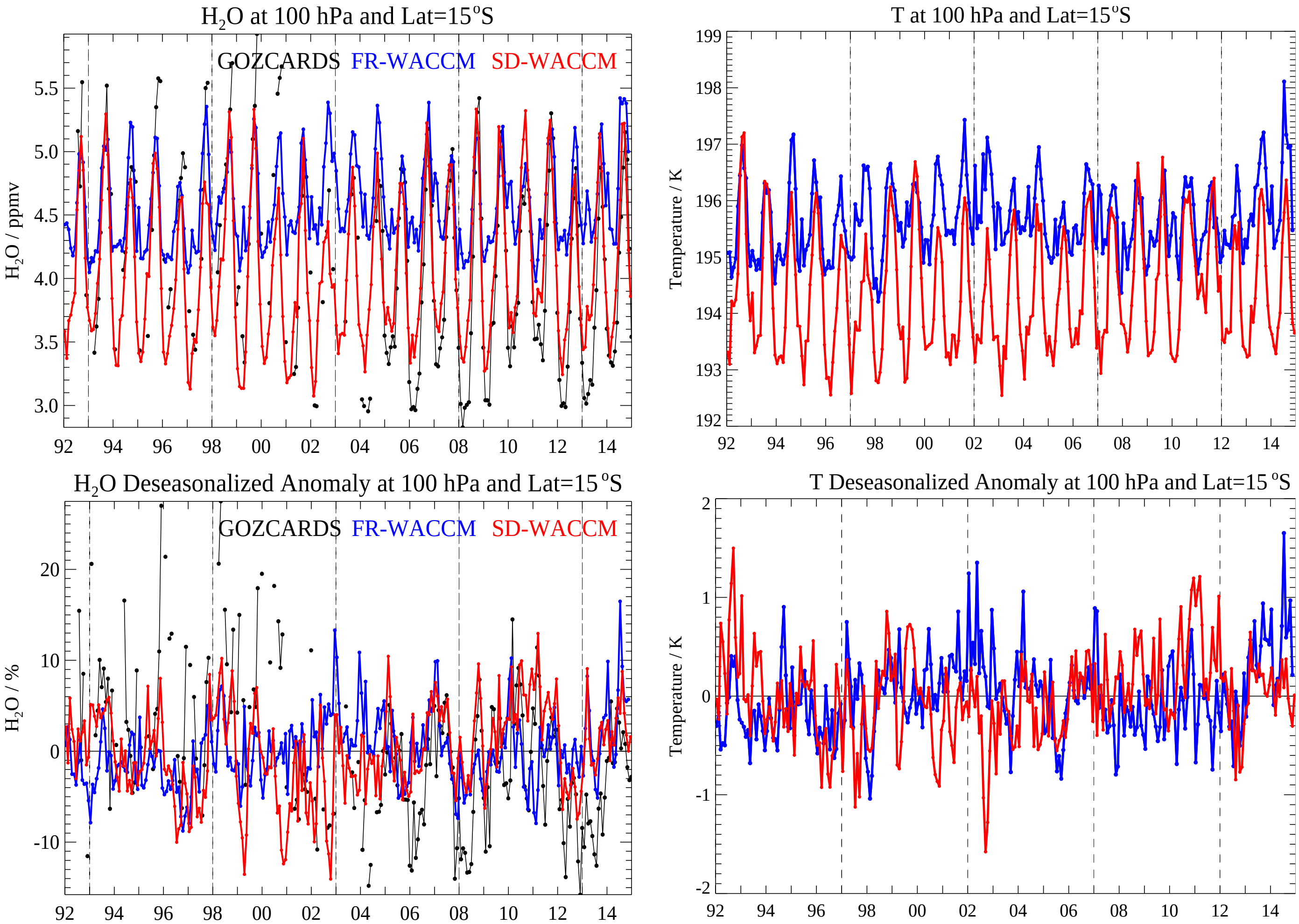

Figure 16. Time series (1992-2014) at 100 hPa and 10°S-20°S for temperature (right two panels) and H_2O (left two panels), with deseasonalized anomalies shown in the bottom two panels. The temperature plots just show the two models (FR-WACCM in blue, SD-WACCM in red), whereas the H_2O series show the comparisons for the models versus GOZCARDS merged H_2O data (in black).

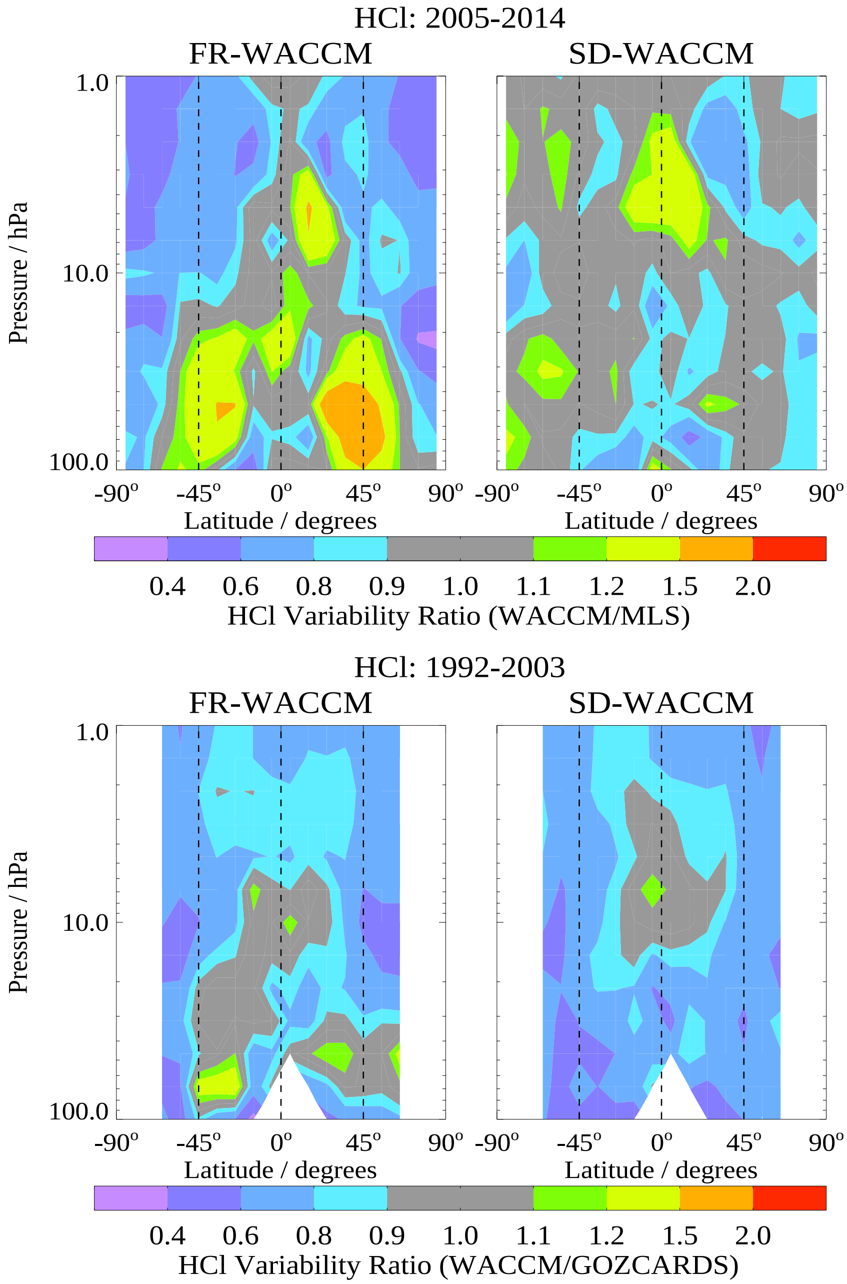

Figure 17. Same as Fig. 14, but for ratios (model/data) of HCl stratospheric variability.

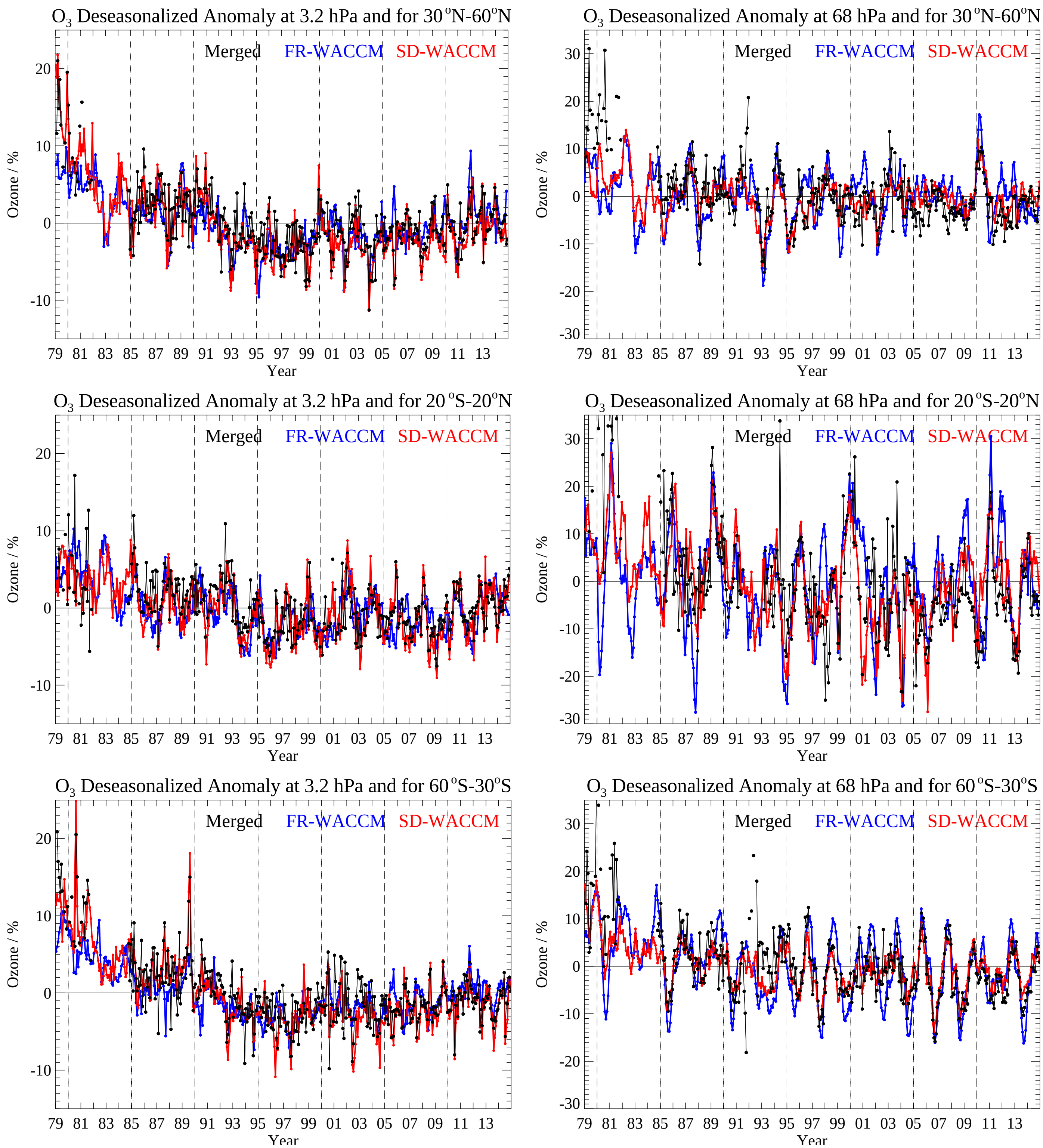

Figure 18. Sample time series of deseasonalized ozone anomalies (%) from 1979 through 2014 from the GOZCARDS data record (version 2.20) compared to the corresponding model anomalies from FR-WACCM (blue) and SD-WACCM (red). Upper stratospheric series at 3.2 hPa are shown in left panels and lower stratospheric series at 68 hPa are on the right; three latitude bins are displayed (30°N-60°N, top; 20°S-20°N, middle, and 30°S-60°S, bottom).

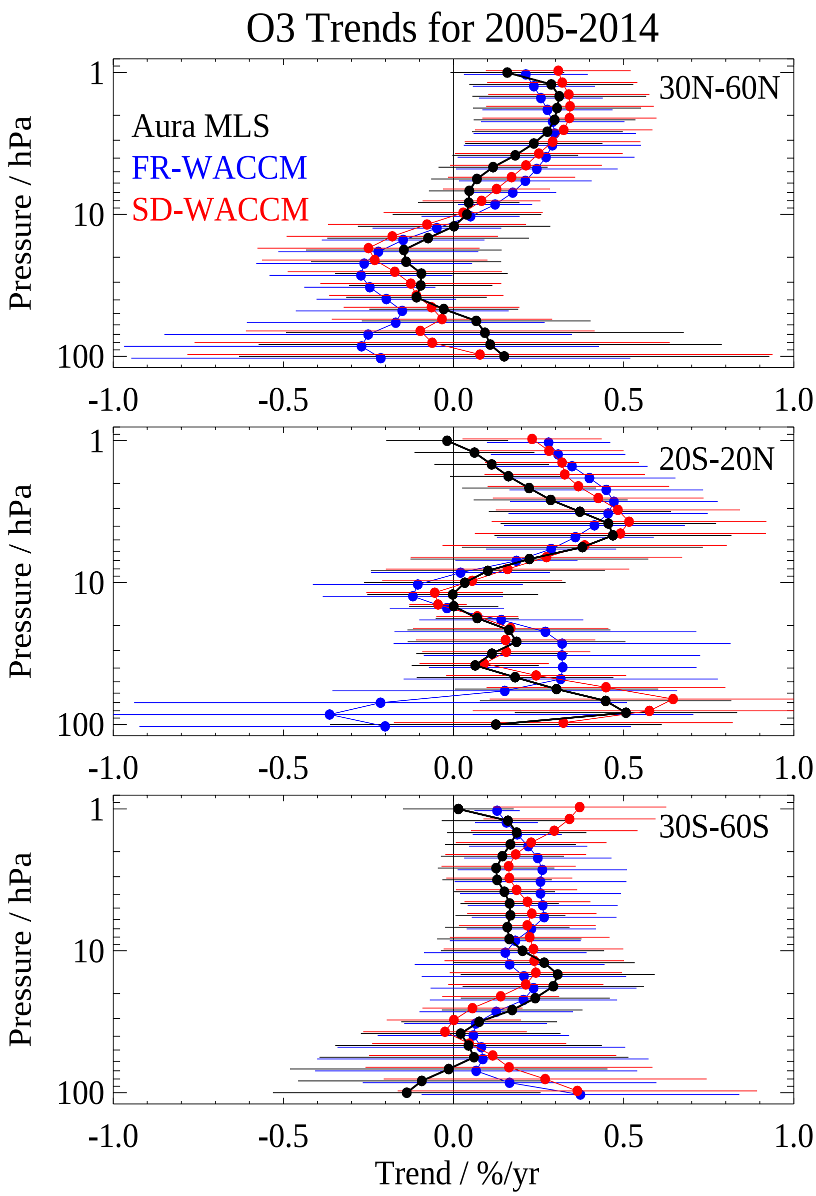

Figure 19. Ozone stratospheric trends for 2005 through 2014 obtained from monthly zonal mean data (version 4.2 Aura MLS) and models (FR-WACCM and SD-WACCM), after multiple linear regression analyses of deseasonalized anomaly time series, as described in the text. Each panel refers to results from different latitude band average series (see legend). The error bars are 2σ estimates based on bootstrap resampling results (see text).

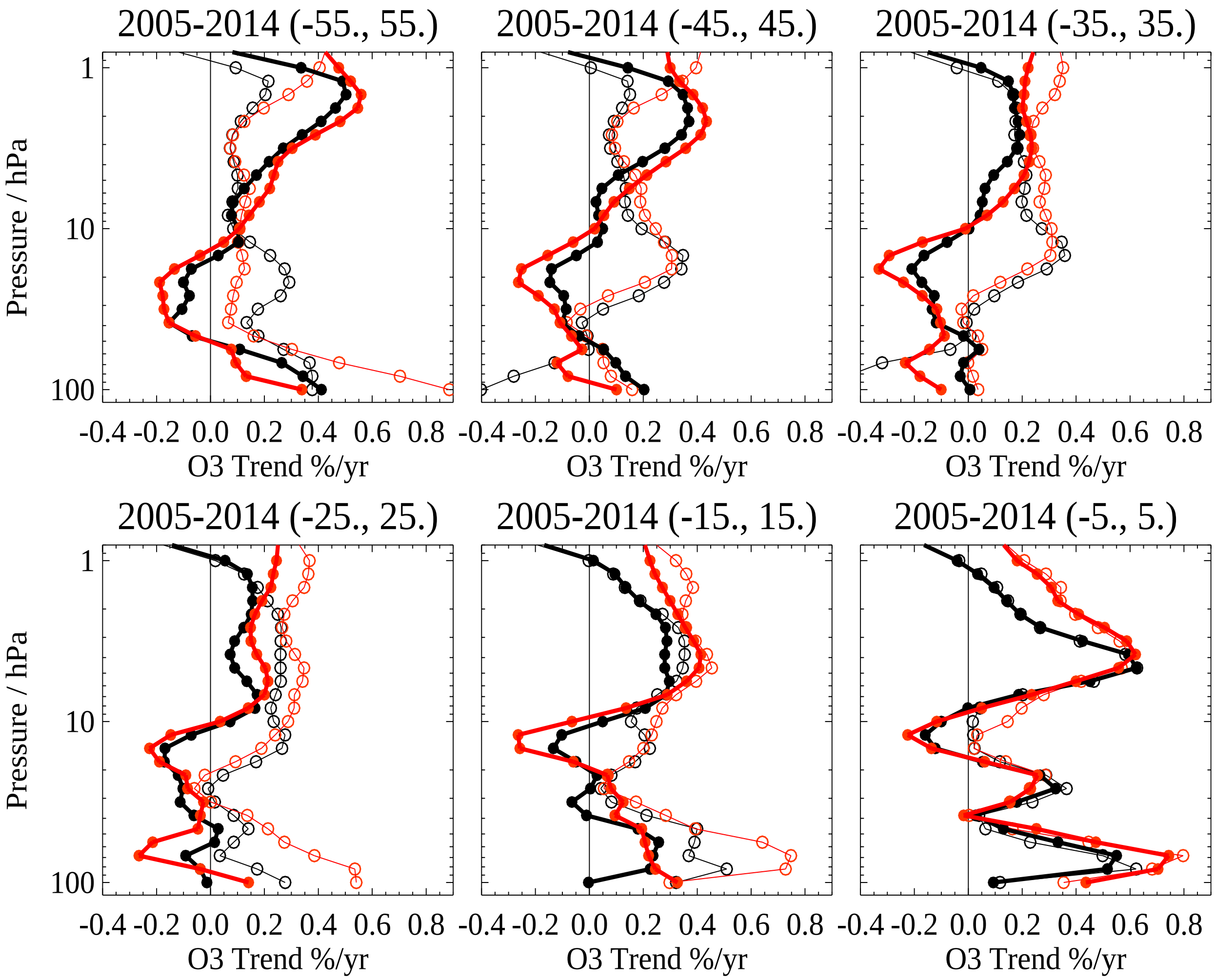

Figure 20. Ozone trends in different latitude bins for SD-WACCM (red) versus MLS data (black) for 2005-2014. Closed and open circles are for northern and southern latitude bins, respectively. For clarity, error bars are omitted here, as these generally show that model/data trend differences are not significant for this time period.

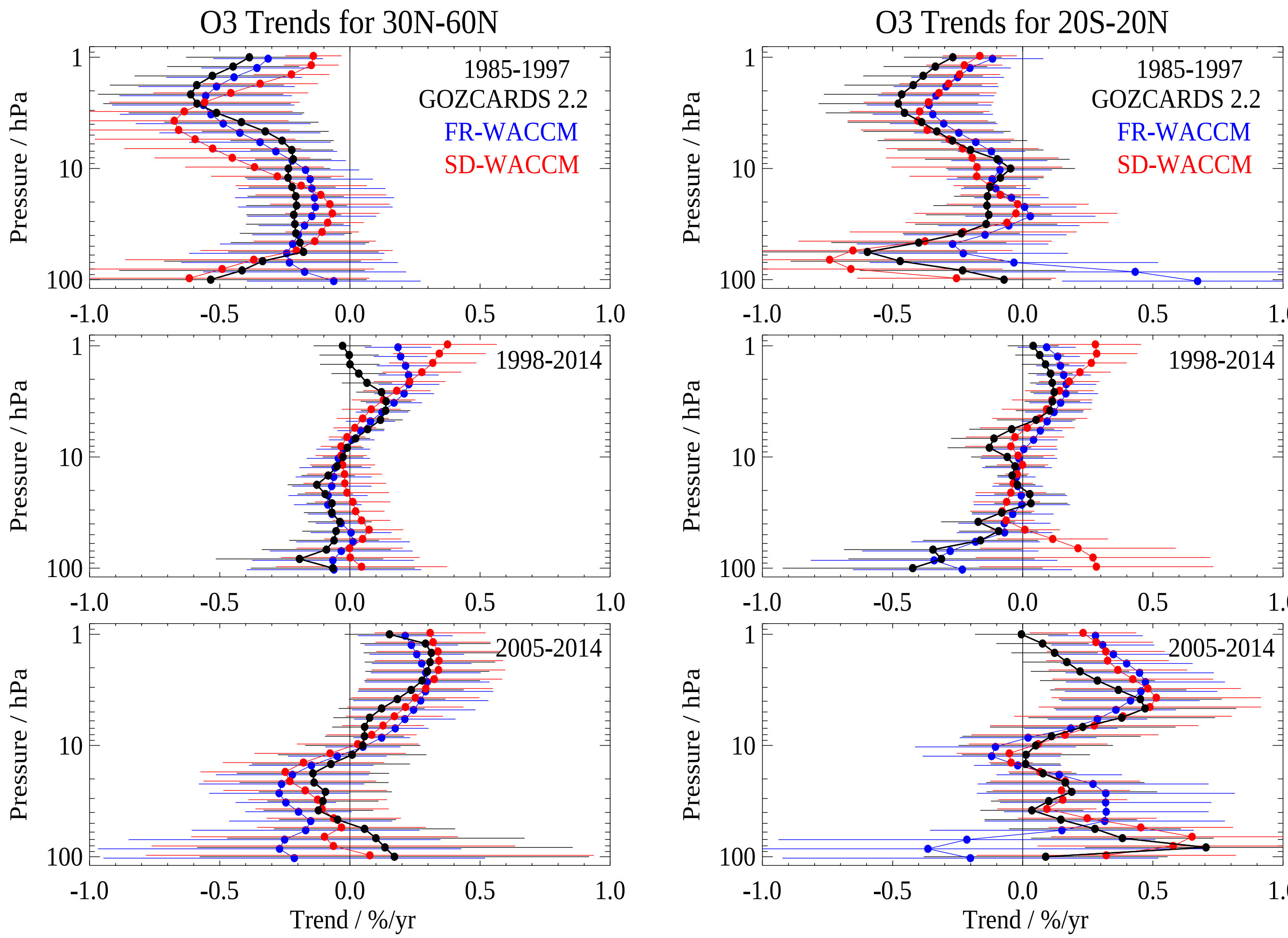

Figure 21. Same as Fig. 19 for ozone trend comparisons, except for the use of two latitude bands (30°N-60°N on left side and 20°S-20°N on right side) and three different time periods (from top to bottom panels, see legend); the data record here is from GOZCARDS merged ozone version 2.20 (see text).

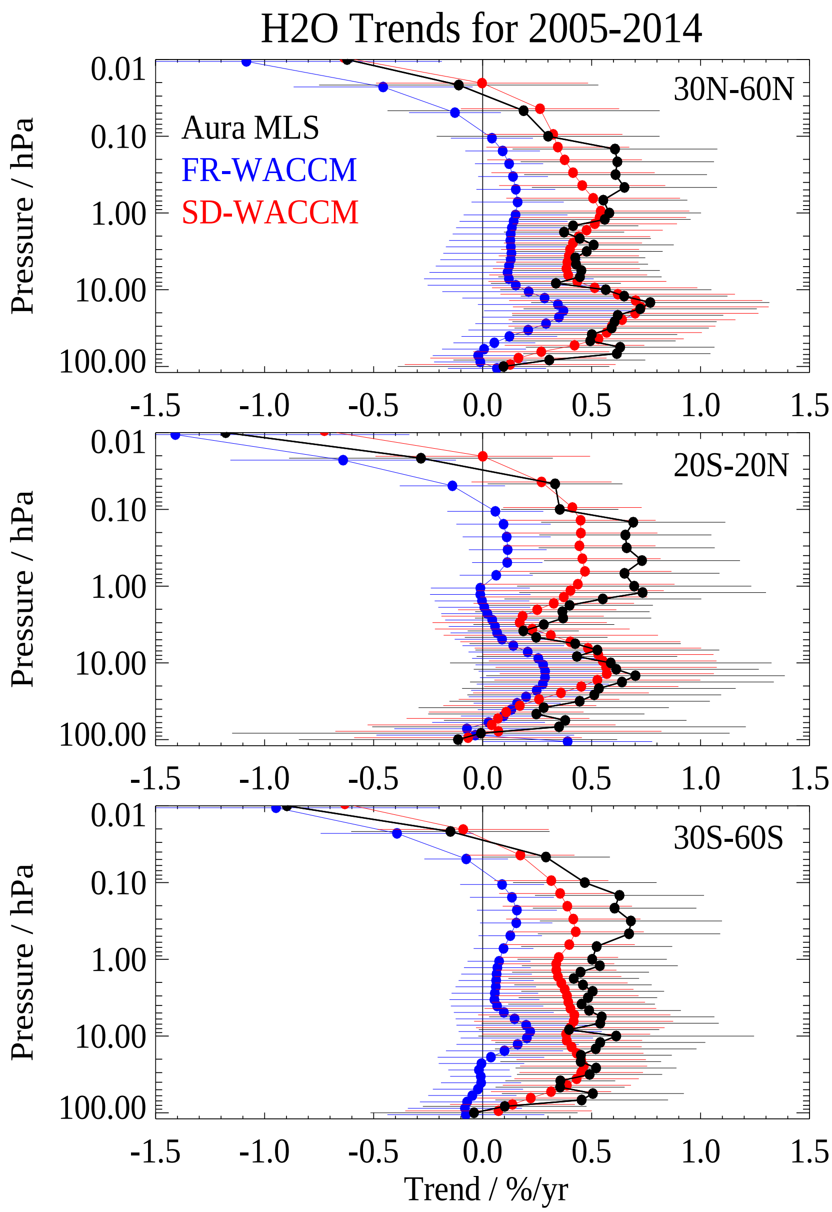

Figure 22. Trends in three latitude bins for stratospheric and mesospheric H_2O from an analysis of the 2005-2014 MLS data and the two WACCM models over the same time period.

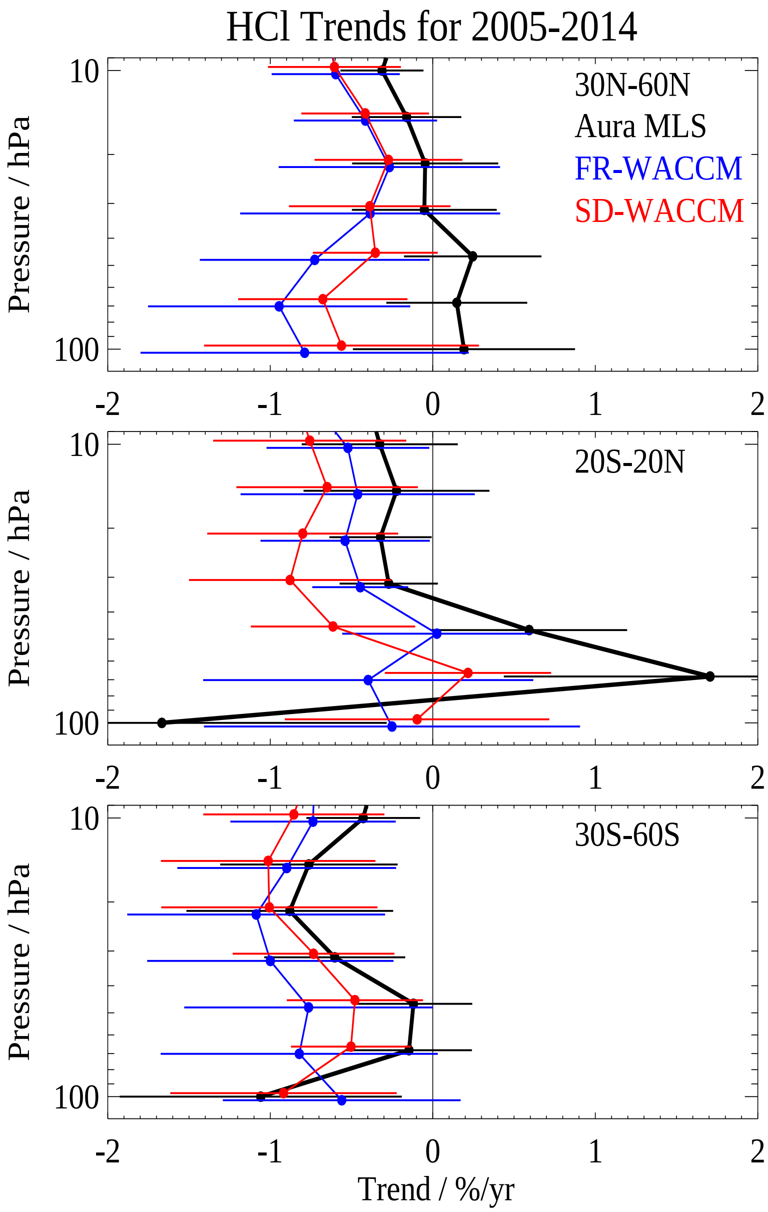

Figure 23. Same as Fig. 19, but for HCl data and model trends in the lower stratosphere.

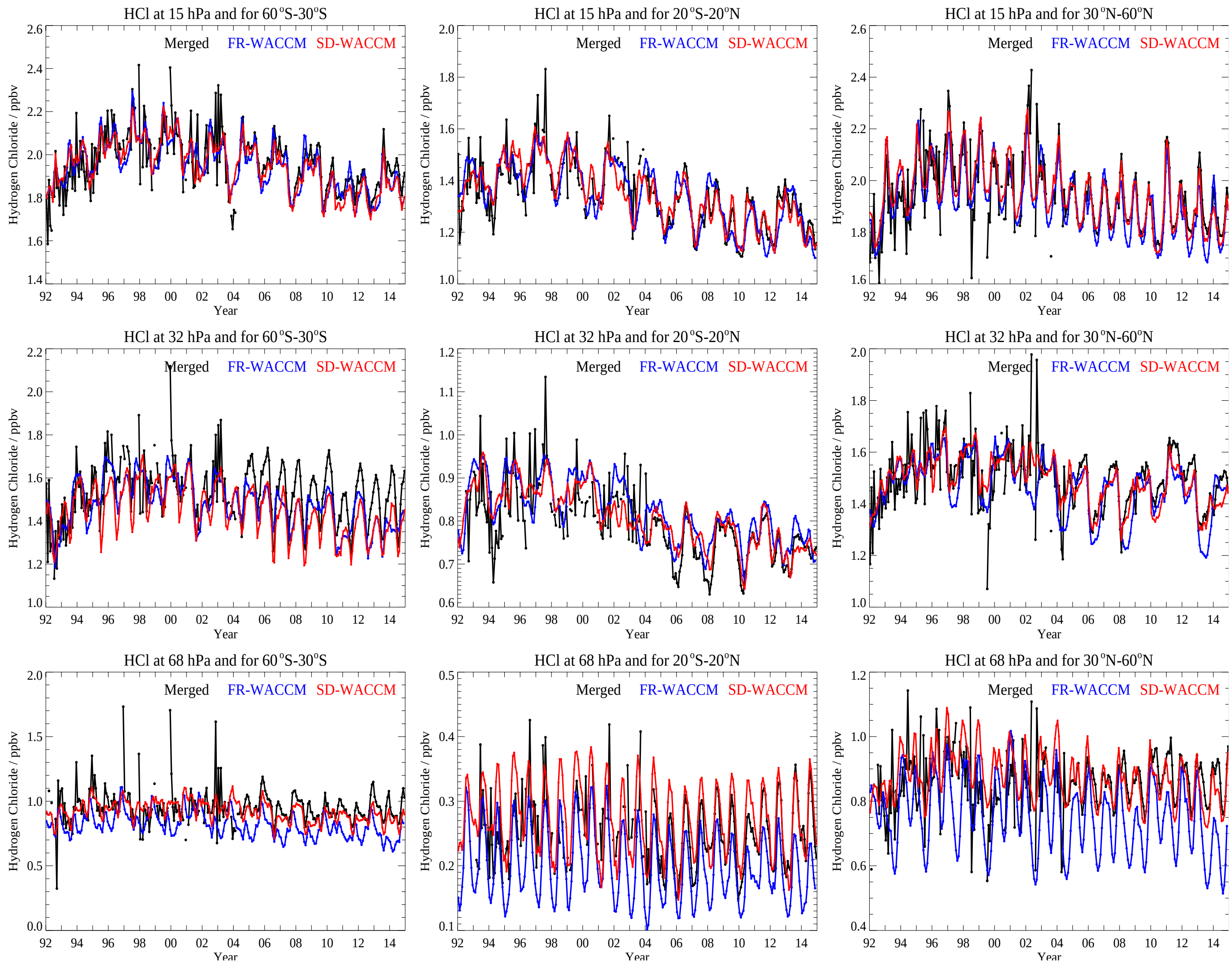

Figure 24. Time series (1992-2014) of lower stratospheric HCl (ppbv) for the GOZCARDS HCl merged data record (black), as well as models (FR-WACCM in blue and SD-WACCM in red). Each panel is for a different pressure level and latitude bin, as labeled (15 hPa, top; 32 hPa, middle; 68 hPa, bottom); the three latitude bins used in this work are 30°S-60°S (left panels), 20°S-20°N (middle panels), and 30°N-60°N (right panels).

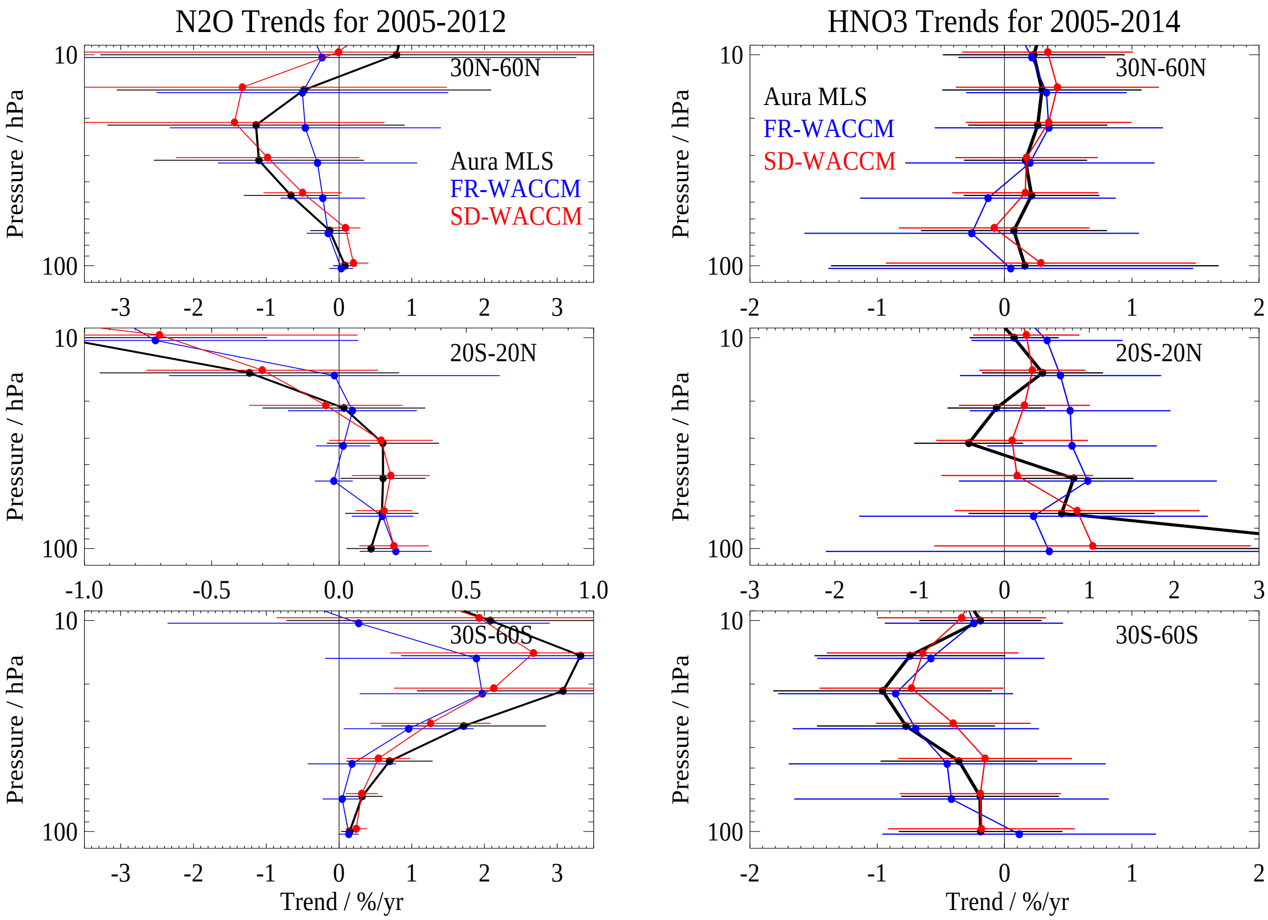

Figure 25. Same as Fig. 23, but for N_2O (left 3 panels) and HNO_3 (right 3 panels) data and model trends in the lower stratosphere. The N_2O data results are from the N2O-640 MLS product (retrieved from the 640 GHz radiometer band data), which was discontinued in 2013 because of an instrument issue affecting this band (see text), and these data and model trends apply to the 2005-2012 period. The HNO_3 trend results (data and models) are for 2005-2014.