We thank Craig for his comments and discussion regarding our manuscript. In addition to the changes we made based on Simon Chabrillat's review, we have added a reference to the 2012 S-RIP proposal (Fujiwara et al., 2012), and the link to the S-RIP website to our manuscript. It is our hope that this paper and these additional references will bring more attention to the importance of intercomparisons and the S-RIP project as a whole.

We can also confirm that we now have the JRA-55 and CFSR datasets available, and have begun using them. As we mention in our conclusions, we will soon extend this study to both of these datasets. Our next study will thus be even more comprehensive, and will focus less on the definitions of the diagnostics, and more on the critical intercomparisons of the most recent reanalyses. We hope to also include GMAO's newest MERRA-2 dataset, which we will have access to soon.

We thank Simon for his detailed comments and very helpful suggestions. Our responses to his comments and questions are:

Major Comments

P. 31366, lines 6-10

The absence of comparison with independent measurements is the main limitation of this study. It can not be justified by the scarcity of such measurements, because many observational datasets include temperature and are not used in data assimilation. Two important examples are

- *the ground-based observations collected by the NDACC network: these include Lidar and ozonesondes (which also measure temperature);*
- *the satellite limb sounders such as UARS-MLS, Envisat-MIPAS or Aura-MLS.*

Several comparisons between such instruments and meteorological analyses can be found in the literature. I think that such comparisons are simply beyond the scope of this paper, but this limitation should nonetheless be mentioned in the Introduction, with proper references to available datasets and published comparisons. I suggest to remind it in the Conclusion as well, because this is an important venue for further research in our field.

We certainly agree that reanalysis comparisons with independent measurements are extremely important when they are possible. In the context of polar processing and the diagnostics that are commonly used, however, there is very little potential for such comparisons. We were unclear in our introduction by making it sound as if there are few independent observations available. What we are really trying to say is that there is a scarcity of independent measurements that are applicable for polar processing diagnostics, which require large-scale temperature and potential vorticity data that currently cannot be provided by satellite or ground-based measurements.

This was a topic of considerable interest at the most recent S-RIP meeting back in September 2014 when several attendees asked if the primary goal of the S-RIP project was to determine the accuracy of reanalyses. From the perspective of those of us who attended the meeting (Zachary, Gloria, and Michelle), it seemed like the consensus was to aim for comparisons with observations when possible, but just as importantly, we should also consider the agreement between reanalyses as valuable information since agreement (or lack thereof) gives users of the data an idea of uncertainty.

We have added a paragraph near the end of our introduction that we hope better explains our reasoning, and the scope of this study.

P. 31367, lines 2-5

If I understand well, the Potential Vorticity from MERRA is interpolated from 42 pressure levels to the model levels, followed by an interpolation to the isentropic levels where all vortex diagnostics are calculated. It looks like significant information could be lost in

this process, while vorticity could be derived from the MERRA wind fields which are distributed directly on the model levels. Please check that the 42 pressure levels have a vertical resolution similar to the model levels, and/or that the diagnostics derived from p-levels PV are sufficiently close to diagnostics derived from model-levels PV. It also looks like the vertical integration of APSC and Avort (to VPSC and Vvort) is done on the vertical grid of isentropic levels (p.31370 line 2). What is its vertical resolution? Using a grid coarser than the model grid could introduce unnecessary errors in the integration. If this is the case, are such errors negligible?

We would have used MERRA PV on the model levels and grid if it were available. Several years ago, we tried to obtain model level PV from GMAO, but unfortunately we were told that it was not archived.

You are correct that two interpolations are performed on the MERRA PV data: one to make the reduced-resolution PV match the model levels and grid, and another to interpolate to isentropic surfaces. Even though this procedure undoubtedly introduces some error, we argue that it is still best to use the PV provided in the MERRA dataset since it does come directly from the model, and the errors introduced in calculating PV from the MERRA winds are likely to be larger than those introduced by the interpolations. We also think that this procedure is what most data users are likely to do, since it is arguably easier to interpolate one field to match the model grid and levels than it is to interpolate all of the model level data to the reduced resolution grid and levels, or calculate PV from MERRA winds. This process of matching the pressure-level PV with the model levels has been used before (e.g., in Manney et al., 2011). We have added this citation in the MERRA portion of the Data & Analysis section. Furthermore, in our experience, the interpolation of MERRA PV preserves the polar vortex edge gradients very well in comparison to GEOS 5.2.0 and 5.9.1 PV data that are, unlike MERRA, provided on the original model grid. Since we use the interpolated PV for vortex diagnostics that primarily depend on well-defined vortex edges, our diagnostics are unlikely to be affected significantly by the interpolated PV. Sentences about this preservation of the vortex edge have been added to the text (also at the end of the MERRA portion of the Data & Analysis section).

The resolution of the isentropic levels we use for the vertical integrations that calculate V_{psc} and V_{vort} is comparable to, but perhaps slightly coarser than, that of MERRA and ERA-I. According to the Knox equation for altitude, the isentropic levels we use (390, 410, 430, 460, 490, 520, 550 and 580 K) are all roughly 1.1 km apart (this information has been added in the text). V_{psc} and V_{vort} are commonly calculated with altitude approximations - some more simple than others. For instance, Rieder and Polvani (2013) calculate V_{psc} from A_{psc} only on two pressure levels (50 and 30 hPa). These diagnostics are not very sensitive to the method used to calculate the volumes, but they are still commonly used because of the strong correlation of V_{psc} (and V_{psc}/V_{vort}) with ozone loss. A sentence about this insensitivity has also been added to the revised paper.

P. 31371, lines 8-16
This description of the calculation of trajectories is not sufficient to allow reproducibility of the results. Livesey (2013) is a simple web page which does not provide the source code, only a brief description and output datasets using (if I understand well) MERRA fields. It looks like Livesey (2013) included diabatic motion, but from section 3.4 we understand that it is not the case here. I think that the explanations on p.31379 lines 15-21 should be transferred here. Even so, some key questions must be addressed: Was LTD code fed with daily wind fields or more frequent analyses, e.g. 6-hourly? In the first case, explain why the errors due to daily update are negligible; in the second case, update the description of downloaded datasets in sections 2.1 and 2.2.

We use the 6-hourly wind and temperature data from MERRA and ERA-I to calculate the trajectories. This information and other clarifications about the trajectory code have been added to the text. We did not update the description of the datasets in sections 2.1 and 2.2, since we mention that 12UT data are used by default “except where specified otherwise” and “unless stated otherwise.”

Section 3.1
*It is explained that in this paper, the word "bias" designates the mean difference between a diagnostic extracted from MERRA and the same diagnostic extracted from ERA-Interim. This is *very* confusing because in the context of analysis evaluation, the evaluation of the "bias" uses an observational dataset as reference and is a proxy for the systematic error present in the analyses. It is explained that the words "relative bias" are an attempt to clarify the concept. This makes the text even more confusing in my view, because in the context of analysis evaluation the "relative bias" is a dimensionless ratio between the absolute bias and some value representing the investigated quantity. The differences discussed in this paper, on the other hand, have the same units as the diagnostics themselves, do not use any independent observations and are not meant to evaluate the validity of either dataset. The typical reader first looks quickly at a paper, reading the titles of the figures and the sections to understand the scope of the paper. This choice of words will unavoidably lead her to believe that some comparison with observations is performed, while this is precisely not the case (see major comment 1 above). Furthermore discussing the units of a "relative bias" is totally counter-intuitive. I strongly recommend to replace throughout the whole manuscript, "bias" and "relative bias" by "mean difference".*

We regret the confusion and clash of terminologies. Our intention was for “relative bias” (in the context of this paper) to mean the “direction” of the diagnostics - in other words, to point out when one dataset is higher/lower in a given diagnostic with respect to another dataset. In hindsight, this was a poor choice of vocabulary due to the standard statistical definitions of “bias” and “relative bias.” We feel, however, that simply replacing these words with “mean difference” does not accurately describe the quantities we use. Therefore, we now refer to

these quantities as monthly Comparison Period Average Differences, or monthly CPADs, to indicate that we're taking monthly means of differences averaged over some period of years (i.e., they're not just mean differences). Hopefully this is more explicit, and much less confusing. We have also removed most mentions of "bias" in the text; only the occurrences of "bias" that were relevant are left in.

P31373, line9; p31374 line 15; p31375 lines 15-16
Why are most figures shown at 580K (fig. 3,4,6,7) while fig. 5 is shown at 490K ? The text mentions repeatedly that several mean diagnostic differences depend on altitude, but this is not shown on any figure. I suggest to show this dependence explicitly for some well-chosen diagnostic, ideally through a vertical profile of this mean difference. For example p31375 lines 15-16: if below 520K there is no convergence towards better agreement, why not show it? It would be interesting to show a case where the disagreement persists even after 2002.

The "Number of Days $T < T_{psc}$" diagnostic is shown at 490 K because it is the level with the greatest number of days below T_{psc} for both datasets, and because it gives an idea of the T_{min} differences at 490 K in addition to those shown at 580 K.

We have added a new figure (Figure 9) to show the average differences of the A_{NAT} diagnostic at all levels up to 580 K over the different comparison periods. We chose to show the altitude dependence for this diagnostic because it fits naturally with the discussion of V_{psc} later in the paper.

Minor Comments
P. 31363, lines 1-12: consider adding some newer references.

The references given in the first few lines are used because they are definitive papers on these topics, but we have added newer references where they are relevant.

P. 31363, line 27: is the word "myriad" really necessary?

It has been removed.

P. 31367, line 13: replace words "In this case", e.g. by "Here"

Changed as suggested.

P. 31368, lines 4-5: for clarity, mention already here the year of introduction of COSMIC GPSRO data in the reanalyses.

Done.

P. 31369, line 10 (also line 24): how is it possible to examine "daily" minimum temperature with only one instantaneous field per day (i.e. at 12:00 UT per sections 2.1 and

2.2)? Please clarify.

We have clarified that these diagnostics are “Daily 12 UT.”

PP.31369-31371: section 2.4 is too long (especially taking into account major comment 3 and next comment suggesting an additional figure). Consider splitting it into "basic" diagnostics (up to P.31370 line 12) and "advanced" diagnostics (VTC, TT195, CT195).

We have split section 2.4 into a “Temperature and Vortex Diagnostics” section, and an “Advanced Dynamical Diagnostics” section at the recommended location.

P. 31370, lines 14-23: The definition of VTC provided here is new. It should be illustrated with a dedicated figure. I recommend "snapshot maps" showing situations with VTC close to 1 and <= 0, for date(s) of special interest (e.g. the initialization dates of the trajectories used for figures 15 and 16). Since it will probably not be possible to distinguish the PV (and temperature) isocontours by both reanalyses, the figure could mention the two numerical values of VTC for each example.

We have added an example figure (now Figure 2) to demonstrate our VTC diagnostic, and included a couple sentences in the text describing this figure.

P. 31370, line 27: vortex split and SSW are two related but different events. Do you mean here simply "vortex split events" ? Same for p.31371 line 26: do you mean e.g. "major SSW with a vortex split" ?

We have clarified p. 3137 line 27 by removing “SSW” and p. 31371 line 26 as suggested.

P. 31373, line 5: provide a reference about this likely impact of AIRS data

We have added references to McNally et al., 2006, and Rienecker et al., 2011. Both of these discuss the large volume of usable measurements from AIRS, and their impact on the assimilation systems.

P. 31373, line 17: "clearly demonstrate" - consider replacing by "clearly show"

Changed as suggested.

P. 31374, lines 9-10: I do not understand "...below either threshold...". Please clarify.

We have clarified this sentence to better explain that we use the ice temperature threshold for the Antarctic because it is a lower (i.e., more sensitive) threshold, which makes the differences clearer for the Antarctic where temperatures often stay below T_{ice} for long periods of time.

P. 31375 line 19: "mixing of air ...and..." -> "mixing of air ...with..."

Changed as suggested.

P. 31376 line 11: Is paper by Livesey et al. already submitted? If no, consider removing this reference as there is already one for this topic; if yes, please update the reference.

The paper by Livesey et al. was submitted recently and should be in press for ACPD very soon. This citation has been updated accordingly.

P. 31376 line 17-20: the blue and red lines on Fig.10 are so close that the differences can not be discussed there in a credible manner. Consider deleting these lines and discussing the differences directly with fig.11.

We feel that these lines are important to demonstrate just how similar the sunlit vortex area diagnostic is from both MERRA and ERA-I. We have modified Figure 10 (now Figure 12), and Figures 6 and 8 for consistency (now Figures 7 and 10), to make the small differences we reference easier to see and understand. In the top panel, ERA-I's line is almost always above MERRA's, which indicates that ERA-I's polar vortex tends to be filled with slightly more sunlight. In the bottom panel, the months May through June show ERA-I's line above MERRA's, while mid-September through October shows MERRA's line above ERA-I's, indicating that the differences change over the season.

P. 31377, lines 8-10: static stability was not discussed in section 2.4. This sentence is not clear and seems not useful to me.

This sentence has been removed.

P. 31378 lines 4-8: These are indeed important caveats on the impact of vertical integration, time averages and smoothing errors. They should be mentioned in the conclusions as well.

We have reiterated this point in bullet 3 of our conclusions.

P. 31379 line 9: "...ERA-I could bias model runs..." -> "...ERA-I could lead model runs..."

Changed as suggested.

P. 31379 lines 15-21: move to end of section 2.4 (see major comment 3 above).

We have left these lines in this section because we think they fit better here, but we have added clarifications to section 2.4.

P. 31381 lines 1-2: It seems to me that very similar results between ERA-I and MERRA can not "lend confidence in transport calculations using winds from these two reanalyses". They simply show that both reanalyses were well constrained by the same datasets. The only way to have confidence in transport calculations is using independent observations (e.g. of chemical tracers).

We have clarified this sentence to say that the similar results lend confidence that using either dataset in transport calculations would give comparable outcomes.

Table 1: I do not understand the difference between "System" (ATOVS, TOVS), "Instrument" which flew on several satellites (e.g. AMSU) and "Satellites" (GOES)

We have removed the “Types” table column. We originally intended for this column to help distinguish what each of the acronyms are (e.g., the name of an instrument versus the name of a satellite), but it is not crucial for the paper.

Figure 2: I guess that the x-axis tickmarks and monthly labels are for the 1st day of each month. If this is the case, please edit the labels to "1Nov", "1Dec" etc. Same for figs. 6,8,10

We have made these changes in addition to those described above, and have changed the figure titles for Figures 6, 8 and 10 (now Figures 7, 10, and 12) for consistency with the next comment.

Figures 3,7,9,11: these figures look very similar and one easily confuses them while reading the text. I suggest to add as figure title (bold font) the name of the diagnostic difference shown on the plot (as for most other figures).

All of these figures (now Figures 4, 8, 11, and 13, respectively) now have titles.

Figure 4: too small, not readable. Please re-arrange the layout (the web page layout of ACP requires wide figures). Legend: please write the three values of sPV used to draw the vortex edge.

We agree that this figure (now Figure 5) is difficult to read in the landscape layout of ACPD, but we think (and hope) it will work much better for the portrait layout of the final ACP publication. The vortex edge sPV values have been added to the figure caption.

Figure 10: please indicate on the legends the units (fraction of hemisphere area)
Figure 11: please indicate on the legends the units (% of hemisphere area?)

We have changed Figure 10 (now Figure 12) to have units of ‘% of a hemisphere’ to be consistent with the units in Figure 11 (now Figure 13), and have included these units in the figure caption. This has also been done for Figure 6 (now Figure 7) to be consistent with Figure 7 (now Figure 8). All text that refers to fraction of a hemisphere has been updated accordingly.

Manuscript prepared for Atmos. Chem. Phys.
with version 2014/09/16 7.15 Copernicus papers of the LaTeX class copernicus.cls.
Date: 19 March 2015

Comparisons of Polar Processing Diagnostics from 34 years of the ERA-Interim and MERRA Reanalyses

Z. D. Lawrence[1], G. L. Manney[2,1], K. Minschwaner[1], M. L. Santee[3], and A. Lambert[3]

[1]New Mexico Institute of Mining and Technology, Socorro, New Mexico
[2]NorthWest Research Associates, Socorro, New Mexico
[3]Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California

Correspondence to: Zachary D. Lawrence (zlawrenc@nmt.edu)

Abstract. We present a comprehensive comparison of polar processing diagnostics derived from the National Aeronautics and Space Administration (NASA) Modern Era Retrospective-analysis for Research and Applications (MERRA) and the European Centre for Medium-Range Weather Forecasts (ECMWF) Interim Reanalysis (ERA-Interim). We use diagnostics that focus on meteorological conditions related to stratospheric chemical ozone loss based on temperatures, polar vortex dynamics, and air parcel trajectories to evaluate the effects these reanalyses might have on polar processing studies. Our results show that the agreement between MERRA and ERA-Interim changes significantly over the 34 years from 1979 through 2013 in both hemispheres, and in many cases improves. By comparing our diagnostics during five time periods when an increasing number of higher quality observations were brought into these reanalyses, we show how changes in the data assimilation systems (DAS) of MERRA and ERA-Interim affected their meteorological data. Many of our stratospheric temperature diagnostics show a convergence toward significantly better agreement, in both hemispheres, after 2001 when Aqua and GOES (Geostationary Operational Environmental Satellite) radiances were introduced into the DAS. Other diagnostics, such as the winter mean volume of air with temperatures below polar stratospheric cloud formation thresholds (V_{PSC}) and some diagnostics of polar vortex size and strength, do not show improved agreement between the two reanalyses in recent years when data inputs into the DAS were more comprehensive. The polar processing diagnostics calculated from MERRA and ERA-Interim agree much better than those calculated from earlier reanalysis datasets. We still, however, see fairly large ~~relative biases~~ differences in many of the diagnostics in years prior to 2002, raising the possibility that the choice of one reanalysis over another could significantly influence the results of polar processing studies. After 2002, we see overall good agreement among the diagnostics, which demonstrates that the ERA-Interim and MERRA

reanalyses are equally appropriate choices for polar processing studies of recent Arctic and Antarctic winters.

1 Introduction

The depletion of stratospheric ozone in the polar regions is a consequence of chemical processing that is strongly dependent upon meteorological conditions (e.g., Solomon, 1999). This polar processing takes place within the stratospheric vortices that form over the Earth's poles in the fall and persist into spring. These polar vortices act as strong barriers to transport and mixing of air across their edges ~~(e.g., Schoeberl et al., 1992; Manney et al., 1994a)~~ (e.g., Schoeberl et al., 1992; Manney et al., 1994a, 2011; Strahan et al., 2013 ; and references therein), providing a pool of isolated air inside them where polar processing can take place (e.g., Schoeberl et al., 1992). The lower stratospheric processes that lead to chemical ozone destruction include the development of polar stratospheric clouds (PSCs), denitrification via sedimentation of PSCs, and conversion of inert chlorine reservoirs to ozone-destroying forms by reactions on the surfaces of PSCs (e.g., Solomon, 1999). Because these phenomena depend critically on temperatures and winds throughout the lower stratosphere (e.g., WMO, 2011, 2015; Brakebusch et al., 2013; Manney et al., 2011; Sinnhuber and references therein), diagnostics related to ozone loss ~~(e.g., WMO, 2007, 2011; Manney et al., 2011, and references therein)~~ requi fields (e.g., winds) and data coverage (e.g., vertically-resolved, hemispheric, multiannual) that cannot be obtained from individual measurement systems such as satellites and radiosonde networks. As a result, the global analyses of meteorological fields provided by data assimilation systems (DAS) that combine many of these measurements are invaluable for polar processing and ozone loss studies. Numerous such DAS analyses are now available, facilitating both observational and modeling studies of polar processing (e.g., WMO, 2011, 2015, and references therein) . However, variations in the representation of meteorological conditions are expected because of differences in the model formulations and ~~resolution~~resolutions, assimilation methods, and assimilated products (Fujiwara et al., 2012) . The existence of these differences raises the possibility of conflicting results and conclusions between similar studies conducted using different DAS analyses.

Polar ozone loss has been the subject of extensive research aimed at quantifying its dependence on ~~myriad~~ dynamical and chemical processes ~~(e.g., WMO, 2011)~~ (e.g., WMO, 2015) . Diagnostics using meteorological conditions to assess the potential for chemical processing, especially PSC formation and chlorine activation, are commonly used. Some of these diagnostics, such as the volume of air below PSC temperature thresholds (V_{PSC}), have been found to have strong links to total column ozone depletion (e.g., Rex et al., 2004; Tilmes et al., 2006; Harris et al., 2010). While some studies have linked changes in V_{PSC} to an expectation of colder winters and greater ozone loss in the Arctic with global climate change (Rex et al., 2004, 2006), others do not support this conclusion (Hitchcock et al., 2009; Pommereau et al., 2013; Rieder and Polvani, 2013). Climate model projections of future ozone loss are also highly uncertain ~~(Charlton-Perez et al., 2010)~~ (e.g., Charlton-Perez et al., 2010) .

Thus, the prediction of future ozone loss is still problematic, and improvements in such predictions will require better understanding of the uncertainties and potential biases in representation of the meteorological conditions upon which polar processing depends so critically in commonly-used DAS.

Previous studies have recognized the importance of understanding the sensitivity of polar processing and ozone loss quantification to different datasets: Davies et al. (2002) showed that two SLIMCAT chemical transport model (CTM) runs driven by horizontal winds and temperatures from the ECMWF (European Centre for Medium-Range Weather Forecasts) and Met Office DAS led to significantly different patterns of denitrification and chlorine activation, and consequently large differences in ozone loss of nearly 20%. Similarly, Santee et al. (2002) found significant discrepancies in PSC formation and composition between model runs that used Met Office temperatures with and without a 3 K reduction. Sinnhuber et al. (2011) found that reducing the temperatures from the ECMWF operational analyses by 1 K in CTM runs for the 2010/2011 Arctic winter resulted in a substantial increase in ozone loss. Brakebusch et al. (2013) reduced GEOS-5 (Goddard Earth Observing System model, version 5) temperatures by 1.5 K in a Whole Atmosphere Community Climate Model simulation of ozone for the 2004/2005 Arctic winter; applying this temperature bias improved the agreement of simulated ozone with measurements from the Aura Microwave Limb Sounder satellite instrument. In other cases, some DAS analyses have been shown to have significant shortcomings for use in polar processing and ozone loss research. For example, Manney et al. (2005b) and Feng et al. (2005) discuss many issues with polar temperatures from the ECMWF 40-year reanalysis (ERA-40), including periods with large spurious vertical oscillations in polar winter temperature profiles (e.g., Simmons et al., 2005). In intercomparisons of temperature diagnostics related to Arctic polar processing of several meteorological analyses, Manney et al. (2003) found that the area with temperatures below PSC thresholds varied by up to 50% between different analyses, and potential PSC lifetimes differed by several days. Manney et al. (2005a) argued that several reanalyses (the datasets referred to therein as the National Centers for Environmental Prediction and National Center for Atmospheric Research (NCEP-NCAR) reanalysis, NCEP–DOE reanalysis-2, and ERA-40) were unsuitable for stratospheric and polar processing studies.

Since the above-mentioned studies, significant advances have been made in modeling and data assimilation, and several additional datasets have become available for constraining the DAS. Long-term reanalysis systems have become much more widely used, and the long-term records of global meteorology based on observational data that they provide are increasingly critical for climate studies. The growing use of reanalysis datasets demands intercomparisons that quantify the differences between them. While numerous intercomparisons have been done (see, e.g., https://reanalyses.org/atmosphere/inter-reanalysis-studies-0), most focus primarily on tropospheric and/or near-surface processes. A few studies have also compared tropical upper tropospheric processes in commonly used reanalyses (e.g., Schoeberl and Dessler, 2011; Fueglistaler et al., 2013). Rieder and Polvani (2013) showed calculations of one polar processing diagnostic, V_{PSC}, from three reanalyses. However, no comprehensive

intercomparisons of diagnostics pertinent to polar processing in the winter lower stratosphere have been done for the reanalyses that are currently in widespread use. In this paper, we present intercomparisons of polar processing diagnostics derived from the National Aeronautics and Space Administration (NASA) Modern Era Retrospective analysis for Research and Applications (MERRA), and the ECMWF Interim Reanalysis (ERA-Interim). These datasets were chosen for this initial study because of their extensive application in numerous stratospheric studies. Rather than focusing on specific seasons and/or a single hemisphere, we present most of our diagnostics for the 1979 – 2013 record of the reanalyses for both Arctic and Antarctic winters. We examine the potential correlation of ~~biases~~ differences between the analyses over the above time period with the timing of changes in observations ingested by their DAS. ~~It is~~

In general, it can be difficult to directly assess the accuracy of reanalyses because there are few ~~, if any, independent measurements~~ independent (i.e., not used in the assimilation) ~~of temperature available~~ measurements that span the full periods of ~~those analyses; the~~ the available reanalysis data. In the context of polar processing, this difficulty is far greater because most of the commonly used polar processing diagnostics require temperature data over large spatial areas and with horizontal resolutions greater than ground-based measurements, such as those from the Network for the Detection of Atmospheric Composition Change (NDACC, see www.ndacc.org), can provide. Large-scale independent temperature measurements can be obtained from some satellite instruments like the Upper Atmosphere Research Satellite Microwave Limb Sounder (MLS), Aura MLS, and Atmospheric Chemistry Experiment Fourier Transform Spectrometer (ACE-FTS), but these measurements typically have biases of their own (e.g., Schwartz et al., 2008; Sica et al., 2008) . Many polar processing diagnostics also require information about the polar vortex from potential vorticity data, which cannot be provided by any measurement system. The degree of agreement between reanalyses is thus an important indicator of their inherent uncertainties and the potential impact of those uncertainties on polar processing studies. Therefore, one of the intentions of this study is to show when the use of multiple reanalyses is recommended to estimate uncertainties of quantities related to polar processing, and when the use of a single reanalysis is sufficient.

In Section 2 we describe the datasets, relevant aspects of the assimilated observations, and the diagnostics and comparison methods we use. The results, presented in Section 3, comprise comparisons of polar processing diagnostics based on temperatures, polar vortex dynamics, and trajectory-based temperature histories. Our conclusions are then summarized and discussed in Section 4.

2 Data and Analysis

2.1 NASA Modern Era Retrospective analysis for Research and Applications

MERRA is a global atmospheric reanalysis that uses version 5.2 of the Goddard Earth Observing System (GEOS) model and assimilation system. It utilizes a combination of 3D-Var assimilation and

Incremental Analysis Update (IAU) (Bloom et al., 1996) to apply corrections from analysis to the forecast model. The MERRA system operates natively on a 0.5° × 0.667° latitude/longitude grid (361 × 540 gridpoints) and uses a hybrid sigma-pressure scheme with 72 vertical levels; the vertical resolution in the lower stratosphere is near 1 km. Further details about the MERRA system are given by Rienecker et al. (2011). The MERRA data files available from NASA's Global Modeling and Assimilation Office (GMAO) are described by Lucchesi (2012). Except where specified otherwise, the MERRA temperature data used here are from instantaneous daily files at 12UT on the model levels and grid. However, the potential vorticity (PV) data from MERRA are available from GMAO only on a reduced 1° × 1.25° latitude/longitude grid (181 × 288 gridpoints) with 42 pressure levels; for the purposes of this study, ~~PV is~~ MERRA PV is linearly interpolated to match the model levels and grid as was done in Manney et al. (2011) . Although these interpolations of the MERRA PV data cause some smoothing in the resulting PV fields, they preserve the strong PV gradients that define the polar vortex edge. All of the PV-based polar processing diagnostics we use depend strongly on the vortex edge and PV gradients (see Section 2.4.1), so these diagnostics are unlikely to be significantly affected by the errors introduced from interpolating MERRA PV to the model grid and levels.

2.2 ECMWF Interim Reanalysis

ERA-Interim (hereinafter ERA-I) is another global atmospheric data assimilation system. The goal of the ERA-I project was to improve upon ECMWF's previous reanalysis, ERA-40, in advance of their planned next-generation reanalysis. It uses 12-hour cycles of 4D-Var assimilation and a T255 spectral model with 60 vertical levels; the vertical resolution in the lower stratosphere is comparable to that of MERRA. The ERA-I system is described in detail by Dee et al. (2011), and the datasets provided by ECMWF from the ERA-I archive are described by Berrisford et al. (2009). ~~In this case~~Here, ERA-I data are used on the highest resolution regular latitude/longitude grid publicly available at 0.75° × 0.75° (241 × 480 gridpoints) on the 60 model vertical levels; this grid has spacing closest to that of the Gaussian grid associated with the spectral model. As with MERRA, unless stated otherwise, the ERA-I temperature and PV data used in this study are instantaneous at 12UT. However, in this case the PV is derived from the provided relative vorticity, temperature, and pressure fields.

2.3 Timelines of Assimilated Observations

Since satellite observations are the primary constraint on reanalysis products at stratospheric levels, it is useful to consider how the data evolve with the introduction of new missions and instruments: Pawson (2012) noted the effect of the TOVS (Tiros Operational Vertical Sounder) to Advanced TOVS (ATOVS) transition in 1998 on middle and upper stratospheric global temperature anomalies from MERRA and ERA-I. At 5 hPa, he shows that ERA-I temperatures dropped suddenly by about 2 K, while in MERRA the mean annual cycle changed noticeably. Pawson asserts that these distinct

discontinuities suggest that more work is needed to properly handle SSU (Stratospheric Sounding Unit) radiances in reanalyses. Fueglistaler et al. (2013) mention that the introduction of COSMIC (Constellation Observing System for Meteorology, Ionosphere, and Climate) GPSRO (Global Positioning Satellite Radio Occultation) temperature data ~~in ERA-I~~ into ERA-I in 2006 caused a temperature shift in the tropics of about 0.5 K at 100 hPa. Simmons et al. (2014) discuss in detail the various effects of the different satellite missions and instruments on ERA-I temperature data, and perform intercomparisons with MERRA and the Japanese 55-year Reanalysis (JRA55).

In many of our diagnostics, we examine how well MERRA and ERA-I agree over the 1979 to 2013 time period. The observations assimilated in the reanalyses change dramatically over this period, and in some cases there are differences between MERRA and ERA-I in the timing of the changes and the observations included. Figure 1 shows a comparison of the primary satellite datasets assimilated in MERRA and ERA-I, compiled from information in Rienecker et al. (2011), Dee et al. (2011) and Simmons et al. (2014). (See Table 1 for the full names of the instruments~~and~~ /satellites listed in Figure 1 and throughout this paper.) The colored regions indicate periods of years between which the data input streams change significantly; in other words, marking times when the differences between MERRA and ERA-I might be expected to shift. In this case, we have chosen boundaries at 1987 (when SSM/I was introduced), 1998 (when ATOVS was introduced), 2002 (inclusion of CHAMP in ERA-I, and AIRS and AMSU-A in both reanalyses), and 2007 (when COSMIC was included in ERA-I).

2.4 Polar Processing Diagnostics and Intercomparisons

~~Many of the~~ The diagnostics we use are ~~described by Manney et al. (2003, 2005a, 2011) , and are~~ designed to assess a wide range of conditions related to polar processing; many of them have been described previously (Manney et al., 2003, 2005a, 2011) . These diagnostics fall into three categories, focusing on assessment of temperatures, vortex characteristics, or air parcel histories. Taken together, the diagnostics used here provide a comprehensive evaluation of the meteorological conditions pertinent to chemical processing and ozone destruction in the polar stratosphere. Because our focus is on assessing the effects of reanalysis differences on studies of polar processing that takes place in the lower stratosphere, we focus in this paper on isentropic levels below about 600 K (approximately 25 km, or 30 hPa).

2.4.1 Temperature and Vortex Diagnostics

The importance of stratospheric temperatures to the formation of PSCs gives rise to the need for temperature diagnostics. Although some recent studies have suggested that liquid PSCs play a dominant dominant role in activating chlorine (e.g., Wegner et al., 2012; Wohltmann et al., 2013), the formation temperatures of solid nitric acid trihydrate (NAT, Hanson and Mauersberger, 1988) and ice particles remain convenient thresholds for the initiation of chlorine activation processes. In this study, we ex-

amine daily 12 UT minimum temperatures, and calculations of area with temperatures below PSC thresholds (henceforth, T_{min} and A_{PSC}, respectively). We also use diagnostics derived from T_{min} and A_{PSC}, such as the number of days during a polar winter with temperatures below PSC thresholds, and the volume of stratospheric air below PSC thresholds (V_{PSC}). For A_{PSC}, vertical temperature profiles of the NAT and ice thresholds are derived using climatological profiles of HNO_3 and H_2O mixing ratios on six per decade pressure levels (Manney et al., 2003), and interpolating to approximately co-located potential temperature surfaces (e.g., the 56.2 and 31.6 hPa levels are referenced to 490 and 580 K, respectively). V_{PSC} is calculated by vertically integrating eight potential temperature levels between 390 and 580 K using the altitude approximation introduced by Knox (1998), which gives altitudes for these levels that are $\sim$1.1 km apart. Altitude approximations are typically used for calculations of V_{PSC}, and these calculations have been shown to be relatively insensitive to the particular approximation used (e.g., Rex et al., 2004; Manney et al., 2011; Rieder and Polvani, 2013).

Since the polar vortex provides the "containment vessel" within which polar chemical processing takes place (e.g., Schoeberl et al., 1992), we also compare diagnostics that characterize vortex strength and size. These include daily 12 UT maximum PV gradients (one measure of vortex strength, e.g., Manney et al., 2011, and references therein), and the area of a hemisphere covered by the vortex (henceforth, MPVG and A_{vort}, respectively). In addition to the total area of the vortex, we also calculate the area of the vortex that receives sunlight each day, since the photochemical processes involved in chlorine catalyzed ozone depletion require sunlight (e.g., Solomon, 1999). Finally, we vertically integrate A_{vort} in the same manner as A_{PSC} to derive the vortex fraction of low temperature air (i.e., V_{PSC}/ V_{vort}). In all cases we use isentropic surfaces, and scale PV into "vorticity units" (s^{-1}) (Dunkerton and Delisi, 1986; Manney et al., 1994b). MPVG is calculated as described by Manney et al. (1994a): scaled PV (sPV) is numerically differentiated with respect to equivalent latitude (i.e., the value of the latitude circle enclosing the same area as a given PV contour); if the maximum gradient occurs at an equivalent latitude poleward of ±80º, we consider the vortex to be undefined and set the maximum gradient equal to zero. To calculate the area of the polar vortex, we use the 1.4×10^{-4} s^{-1} sPV contour as a simple proxy for the vortex edge (e.g., Manney et al., 2007). The total area of the vortex is then the area of the contour. The sunlit area is the area inside the vortex-edge contour that is equatorward of the daily polar night latitude at 12UT.

2.4.2 Advanced Dynamical Diagnostics

One of our diagnostics is best described as a hybrid temperature-vortex diagnostic. It is the concentricity of the polar vortex with regions of temperatures below the NAT PSC threshold (henceforth referred to as vortex-temperature concentricity, or VTC). VTC is adapted from the concept of concentricity as discussed by Mann et al. (2002). We calculate it using the simple formula

$$\mathrm{VTC} = 1 - \frac{\mathrm{GCDist(Vortex\ Centroid,\ Cold\ Region\ Centroid)}}{\mathrm{GCDist(Pole,\ Equiv.\ Lat.\ of\ Vortex\ Edge)}} \quad (1)$$

where GCDist(x,y) is the great circle distance between x and y. This definition provides an intuitive picture of the vortex/temperature relationships under extreme conditions: a maximum value of 1 for collocated centroids (completely concentric), and values less than or equal to 0 for a cold region centroid approximately at or outside the vortex edge. Centroid locations are calculated as described by Mitchell et al. (2011) and Seviour et al. (2013), with the 1.4×10^{-4} s^{-1} sPV and pressure-dependent NAT PSC temperatures as the edge values for the vortex and cold regions, respectively. For simplicity, we only calculated one centroid for each field. This means that, for example, during vortex-split ~~sudden stratospheric warming (SSW) events, daughter~~ events, offspring vortices were not characterized individually. Under conditions where multiple closed contours exist, VTC values may be calculated using centroids that lie completely outside of the regions of interest. While it would be important to more accurately characterize split vortices for detailed dynamical studies, this simplification does not significantly affect the broad climatological comparisons we are focusing on here. Figure 2 shows an example of our VTC diagnostic using maps of MERRA PV at 490 K with the centroids of the polar vortex (the black squares) and cold regions (magenta X marks) overlaid. The dates shown were chosen for their extremes of VTC - one nearly concentric case (7 Feb 1996) when both MERRA and ERA-I have VTC close to 1, and one non-concentric case (28 Feb 1996) when MERRA and ERA-I have VTC close to 0.

We also use trajectory diagnostics to examine temperature histories of air parcels; these provide important information about the potential for polar processing. Here we use a trajectory code adapted from the Lagrangian Trajectory Diagnostic (LTD) code described by ~~Livesey (2013)~~ Livesey et al. (2015) , which advects parcels using ~~a~~ fourth-order Runge-Kutta ~~method~~integration. The code uses linear interpolation to approximate winds and other fields at intermediate timesteps, and for determining the values of these fields at parcel locations. The trajectories are calculated using the 6-hourly (00, 06, 12, and 18 UT) wind and temperature fields from MERRA and ERA-I. Our standard runs consist of isentropic 15-day forward and backward (30 days total) trajectories of parcels initialized at 00UT ~~integrated with 15-minute timesteps for parcels~~ on an equal-area grid. For this paper, we ~~initialized parcels~~ used 15-minute timesteps for the Runge-Kutta integration and parcels initialized on the 490 K potential temperature surface configured on an equal-area grid poleward of ±40° latitude with 0.5° $\times$ 0.5° equatorial spacing (corresponding to ~30,000 parcels). We examine ~~the~~ parcels initialized in cold regions defined by T $\leq$ 195 K (the approximate NAT threshold at 490 K) ~~,~~and the amount of time ~~they spend below the threshold~~ these parcels spend below 195 K before and after the initialization date. From this subset of parcels, we calculate and compare distributions for total time spent below 195 K (TT195) and continuous time spent below 195 K (CT195) as described by Manney et al. (2003, 2005a).

3 Results

3.1 Monthly ~~Relative Biases~~Comparison Period Average Differences

Figure ~~2~~ 3 shows an example of the type of differences we calculate for most of the diagnostic intercomparisons (in this case T_{min}), for 1979 through 2013. The orange line showing the differences for 2012/2013 provides an example of the magnitude of differences in an individual recent year. Note that there are large day-to-day variations in differences, and that the differences are largest in December 2012 and early January 2013, when the polar vortex was unusually disturbed prior to/during a major ~~vortex-split SSW(?)~~ sudden stratospheric warming (SSW) with a vortex split (Coy and Pawson, 2015) . The departure of the daily average differences (thick black line) from zero suggests a persistent ~~relative bias~~ difference between the two reanalyses that dominates for much of the 34 years. Since ~~this bias varies~~ these differences vary over the season, we define ~~monthly relative biases~~ "monthly comparison period average differences" (monthly CPADs) between the datasets as the monthly ~~averaged mean~~ means of the average daily differences over the ~~years in each of the periods~~ comparison period years defined in Figure 1. These quantities provide a compact means of summarizing the agreement between the datasets for a given month ~~and~~ within a subset of years, and of comparing the magnitude of differences between comparison periods for a particular diagnostic. We ~~refer to these as "relative" biases to emphasize that they represent biases~~ emphasize that monthly CPADs are meant to diagnose significant differences between the two datasets~~, as opposed to~~ ; they are not an assessment of the absolute accuracy of either. For the rest of this paper, we use the convention of subtracting ERA-I from MERRA (that is, MERRA minus ERA-I) to calculate ~~monthly relative biases. As such~~the monthly comparison period average differences. Thus, differences in a diagnostic greater (less) than zero indicate that, on average, MERRA is greater (less) than ERA-I.

3.2 Temperature Diagnostic Intercomparisons

The seasonal progression of polar minimum temperatures provides an indication of when conditions favor the development of PSCs. The dependence of PSC formation on a temperature threshold implies that conclusions drawn from the T_{min} diagnostic are most sensitive to differences at the beginning and ends of the season when minimum temperatures first drop below or rise above PSC thresholds. For the Arctic, these periods are typically around the beginning of December and mid-March, respectively, but large interannual variability and the common occurrence of mid-winter SSWs can result in much earlier or later threshold dates in individual years (Manney et al., 2005a, and references therein). In the Antarctic, the threshold periods tend to be in the first half of May and in mid-October. Figure ~~3~~ 4 shows the monthly ~~relative biases~~ CPADs between MERRA and ERA-I at the 580 K level ($\sim$30 hPa, corresponding to an approximate T_{NAT} of 193 K). For years preceding 2002, MERRA consistently has lower minimum temperatures than ERA-I. The Antarctic ~~relative biases~~ monthly CPADs in this period are particularly large, with differences between -5 ~~to~~ and -6 K

in some months, in contrast to the largest differences in the Arctic, which are about -1.4 K. Differences in 1998 – 2001, after the introduction of the ATOVS instruments, but before the introduction of the Aqua instruments, are significantly reduced over those prior to 1998. In both hemispheres, there is a distinct shift in agreement after the introduction of the Aqua instruments from 2002 onward (likely due to the vast increase in the number of observations included by assimilating AIRS data - see, e.g., McNally et al. (2006) and Rienecker et al. (2011)). This is especially easy to see for the Antarctic, where the ~~relative biases~~ differences are reduced to values akin to those in the Arctic. In the Northern Hemisphere (NH), the shift marks the first period in which ERA-I minimum temperatures become consistently lower than those from MERRA. At lower levels down to about 460 K (not shown), the T_{min} ~~relative biases~~ monthly CPADs are smaller in magnitude (e.g., at 490 K they are between -0.4 and 0.7 K in the Arctic, and -2 to 2 K in the Antarctic), and have different seasonal variations, with ERA-I having comparatively more months in the first three time periods with lower minimum temperatures.

To examine the global variation of temperature differences between MERRA and ERA-I, Figure ~~4~~ 5 shows maps of mean temperature differences (averaged over the time periods of interest from Figure 1) for the months with the largest ~~relative~~ T_{min} ~~biases in Figure 3.~~ monthly CPADs in Figure 4. The maps for the Arctic, with the more symmetric color bar, clearly ~~demonstrate~~ show that the regions where MERRA is colder tend to be mostly confined between ±90° longitude and poleward of 60° latitude. This is a preferred direction for the polar vortex and cold region to be shifted off the pole in Arctic winters (e.g., Waugh and Randel, 1999). Outside of this area, ERA-I temperatures are, for the most part, lower than those from MERRA. In the Southern Hemisphere (SH), the first three time periods show that the regions that are colder in MERRA are fairly symmetric poleward of the 60° S latitude circle. Examination of the mean temperature fields (not shown) for these periods suggests that larger ~~biases~~ differences are associated with lower temperatures. The maps for the following two time periods, from 2002 to 2013, demonstrate that the change seen in Figure ~~3 reflects a distinct~~ 4 reflects a large shift throughout the polar regions: the largest temperature differences during these periods are confined to relatively small regions, and overall the temperature differences lie between ±0.4 K in both hemispheres. Simmons et al. (2014) show extratropical zonal-mean temperature differences between ERA-I and MERRA at 30 hPa that indicate that the extratropics (± 20 – 90° latitude) are colder in ERA-I than in MERRA for most of the reanalysis period. The results shown here are generally consistent with that finding, but suggest that this result does not always hold true poleward of ±60° latitude in winter/spring.

The number of days below PSC thresholds, a diagnostic derived from T_{min}, for winters in the 1979 through 2013 period is shown at 490 K (∼56 hPa) for both hemispheres in Figure ~~5; we~~ 6. We show days below T_{ice}, rather than T_{NAT}, in the Antarctic because the ~~period~~ periods with temperatures below ~~either threshold is much longer~~ T_{NAT} and T_{ice} are much longer than the periods below these temperature thresholds in the Arctic, making the differences ~~clearer~~ in the SH more

apparent for the lower (i.e., more sensitive) ice threshold. This diagnostic indicates the approximate duration of the period with conditions conducive to polar processing. Overall, the number of cold days from ERA-I is greater than the number from MERRA in both hemispheres at this level. The Antarctic differences before 2002 are quite large, with some years showing ERA-I having over 10 more days with temperatures below the ice PSC threshold than MERRA. At levels up to 580 K (not shown), the situation in the Antarctic is opposite; that is, ~~there are~~ MERRA has significantly more days (sometimes between 20 and 30 days in years before 2002) with temperatures below the ice PSC threshold ~~in MERRA than in~~ than does ERA-I. These results, along with those discussed ~~about~~ above for T_{min}, suggest that not only may the choice of dataset have a large influence on analysis and modeling of polar processes in the Antarctic for years preceding 2002, but also that effects may vary qualitatively and quantitatively in the vertical. In contrast to those for the Antarctic, the ~~magnitude~~ magnitudes of differences in the Arctic at higher levels up to 580 K are largely similar to those at 490 K, but with MERRA having more cold days than ERA-I.

Average values for the total area with temperatures below the NAT threshold at 580 K are displayed in Figure ~~6.~~ 7. The Arctic maximum mean area is just above 3% of the hemisphere for ERA-I, while MERRA reaches nearly 4%. In a similar fashion, the Antarctic maximum mean area is about 11% of the hemisphere for ERA-I, while that for MERRA reaches nearly 12%. Differences in these mean values between the reanalyses vary considerably with the season in both hemispheres. Figure ~~7~~ 8 shows the corresponding monthly ~~relative biases~~ CPADs for A_{NAT}. Here it is worth noting that the ~~low biases~~ small differences in the last month of the season in each hemisphere are expected as these include many days that have zero ~~biases because~~ differences since minimum temperatures have risen above the NAT threshold in both reanalyses. Consistent with the T_{min} ~~relative biases~~monthly CPADs, there is much closer agreement after the first three periods. The relatively frequent occurrence of warm Arctic winters after 1998 (e.g., Manney et al., 2005a) suggests that the third ~~Northern Hemisphere period~~ comparison period for the NH might be less easily compared with the ~~previous period with unusually cold Arctic winters (e.g., Pawson and Naujokat, 1999)~~ second comparison period (which contained unusually cold Arctic winters, e.g., Pawson and Naujokat, 1999), but the marked decreases of ~~relative biases~~ the monthly CPADs in the SH still indicate a substantial effect due to changes in assimilated observations. In most cases the ~~relative biases~~ monthly CPADs are positive, indicating that MERRA tends to have larger cold regions than ERA-I at this level in both hemispheres. For levels below 580 K, down into the upper troposphere/lower stratosphere (UTLS) region at 390 K~~(not shown)~~, MERRA still tends to have larger cold regions than ERA-I, but the conditions are different: the ~~relative biases~~ differences are much smaller in all periods, generally lying between -1 and 1% of a hemisphere in the Antarctic, and -0.4 and 0.4% of a hemisphere in the Arctic. ~~Although the relative biases are smaller~~ Figure 9 demonstrates this behavior; it shows time series contour plots of the MERRA minus ERA-I A_{NAT} differences averaged over the comparison periods from Figure 1. Although the differences are smaller at lower levels, the same convergence towards

better agreement we see at 580 K is not always seen at levels below 520 K, especially in the Arctic around 410 and 430 K.

3.3 Vortex Diagnostic Intercomparisons

Maximum PV gradients (MPVG) indicate the strength of the polar vortex as a barrier to transport and mixing of air from lower latitudes ~~and~~ with the cold vortex air where chlorine activation takes place (e.g., Manney et al., 2011). The seasonal evolution and mean values of maximum PV gradients at 490 K are shown in Figure ~~8.~~ 10. The primary difference between the hemispheres is seen in the average values; the Arctic maximum gradients tend to level off at around 10^{-5} $\mathrm{s}^{-1}\mathrm{deg}^{-1}$ early in the season, while the Antarctic maximum gradients steadily increase up to approximately 1.7 $\times 10^{-5}$ $\mathrm{s}^{-1}\mathrm{deg}^{-1}$ near the end of the season. The MPVG ~~relative biases~~ monthly CPADs are shown in Figure ~~9.~~ 11. Note that the scaling of the values is 10^{-6} $\mathrm{s}^{-1}\mathrm{deg}^{-1}$. This means ~~a relative bias~~ that a difference of 1 would be 1/5th the height of a grid-box from Figure ~~8.~~ 10. As such, the ~~relative biases~~ monthly CPADs are overall quite small, but the monthly and comparison period variations still provide useful information. For instance, the greater MERRA MPVGs in the months when the vortex usually weakens (March or April in the Arctic, November in the Antarctic) suggest that the polar vortices as represented by ERA-I tend to weaken earlier than those in MERRA. Furthermore, a shift in agreement contemporaneous with those seen in Figures ~~3 and 7~~ 4 and 8 is also present. In this case, however, the ~~relative biases~~ monthly CPADs increase considerably for the ~~Northern Hemisphere~~ NH rather than decrease. This behavior, with agreement in the SH (NH) improving (degrading) slightly, is also seen at other vertical levels between 460 and 580 K (not shown).

Similar to A_{PSC}, the sunlit area of the vortex provides an approximate quantitative measure of the area on a given vertical level where chlorine-catalyzed ozone destruction can take place. Different sunlit area diagnostics have been used in detailed polar processing studies to establish correlations between the coverage (e.g., Feng et al., 2007) and duration ~~(e.g., Rex et al., 1999; **?**, in preparation)~~ (e.g., Rex et al., 1999; Livesey et sunlight and ozone loss; these diagnostics are usually somewhat computationally intensive, whereas the diagnostic used here (described in Section 2.4 above) is simple enough to compute for multiple long-term datasets. Figure ~~10~~ 12 shows time series averages and ranges of the sunlit vortex area diagnostic as a ~~fraction~~ percentage of a hemisphere at 490 K. The seasonal variations of the sunlit vortex are very similar in both datasets. ~~There are~~Small differences can, however, ~~small biases that can~~ be seen: the ERA-I Arctic polar vortex tends to be filled with slightly more sunlight than the MERRA vortex, while in the Antarctic, the ~~bias changes~~ differences change over the season. Consistent with this, the sunlit vortex ~~monthly relative biases in Figure 11~~ area monthly CPADs in Figure 13 show predominantly negative values in the Arctic (~~indicating the bias towards ERA-I~~ in agreement with Figure 12, which showed ERA-I values greater than those from MERRA), and ~~biases~~ differences that change sign in the Antarctic. Like the NH MPVG ~~relative biases~~monthly CPADs, the NH sunlit vortex ~~relative biases~~ area monthly CPADs increase to a maximum by the

~~5th~~ last time period (2007-2013) of Figure 1. The overall maximum ~~relative biases~~ monthly CPADs occur in November and February of this final time period, both months when a substantial portion of the vortex is typically in darkness, and when varying sunlight may affect chlorine activation whenever temperatures are low enough for PSC formation. In contrast to those for the NH, the SH ~~relative biases~~ monthly CPADs improve slightly over the reanalysis period. ~~Biases~~ Differences of the small magnitude shown here are unlikely to have large effects on polar processing in either hemisphere, especially in the Antarctic, where the vortex is larger, colder, and less variable from year to year than in the Arctic. The vertical structure of the sunlit ~~area biases~~ vortex area differences (not shown) is complicated, but for recent years the ERA-I NH polar vortex tends to have a larger ~~sunlit area~~ area in sunlight than that in MERRA at levels between 460 and 550 K, while the opposite is true for the SH. The ~~biases~~ differences in sunlit vortex area shown here primarily arise from ~~similar biases~~ corresponding differences between MERRA and ERA-I in total vortex area (A_{vort}, not shown). However, small ~~differences between the relative biases~~ discrepancies between the monthly CPADs of sunlit and total vortex area indicate ~~a very small effect from~~ that differing vortex positions ~~. Examination of the relative vorticity and static stability from MERRA and ERA-I suggests a small, but persistent, global bias towards larger magnitude static stability values in ERA-I, which likely contributes to the biases in vortex size~~in the two reanalyses also affect the monthly CPADs slightly.

The winter mean vortex fraction of cold air diagnostic helps to identify years with conditions favorable for PSC formation and ozone loss. In the Arctic, years with low values of V_{PSC}/V_{vort} correspond to the winters with very short cold periods usually associated with a disturbed vortex and mid-winter SSWs. Because of its relationship to column ozone loss, V_{PSC} has been used extensively in climatological ozone loss studies ~~(e.g., Pommereau et al., 2013; Rex et al., 2004, 2006; Rieder and Polvani, 2013; Tilmes et al., 2006)~~ V_{vort} as a stand-alone diagnostic is much less common; however, the volume of air within the vortex is an indicator of the absolute size of the region over which chemical ozone loss can occur. Figures ~~12 and 13~~ 14 and 15 show winter mean V_{PSC}/V_{vort} for the Arctic and Antarctic, respectively, along with the corresponding winter mean V_{PSC} and V_{vort} values separately. In general, MERRA tends to have a larger volume of cold air in both hemispheres, consistent with the larger A_{NAT} ~~relative biases~~ monthly CPADs shown in Figure ~~7.~~ 8 and the average differences shown in Figure 9. The only major exceptions are the years from 1979 to 1987 in the Antarctic, where ERA-I V_{ice} is a bit larger than that in MERRA. The lack of improvement in agreement for V_{PSC} follows directly from the prior discussion of A_{NAT}; since the agreement of A_{NAT} between MERRA and ERA-I below 520 K generally does not improve much over time, neither does that of V_{PSC}. Consistent with the ~~biases~~ differences discussed in the sunlit vortex area (Figures ~~10 and 11~~12 and 13), MERRA's V_{vort} values are very similar to those of ERA-I. Thus, on average, the differences in V_{PSC}/V_{vort} between MERRA and ERA-I reflect the ~~biases~~ differences seen in V_{PSC}. Exceptions to this are in 2007-2013 in the NH, when slightly smaller MERRA V_{vort} values emphasize the ~0.01 larger V_{PSC}/V_{vort} values seen in MERRA, and in 1979-1987 in the SH, when the ~0.02 larger ERA-I V_{PSC}/V_{vort} values represent

larger cold fractions of smaller vortices. This analysis demonstrates several caveats for comparisons of diagnostics that rely on vertical integration, time averages, or combinations of both. Their dependence on time and/or altitude can lead to cancellations and smoothing when integrated and/or time-averaged, making comparisons of the final results difficult to interpret since agreement (or lack thereof) can come about for the wrong reasons. In this case, the agreement of winter mean V_{PSC}, V_{vort}, and V_{PSC}/V_{vort} between the reanalyses was affected by the representation of several conditions. Because the ~~relative biases~~ differences between MERRA and ERA-I in A_{PSC} and A_{vort} ~~change differently with~~ are not uniform in time and altitude (see Figure 9), better agreement at some levels does not necessarily correlate with better confidence in our knowledge of V_{PSC}, and consequently V_{PSC}/V_{vort}. This suggests that these diagnostics function better as qualitative measures, and argues for considerable caution in interpretation of their time variations or trends.

The concentricity of the vortex with regions of air with $T \leq T_{NAT}$ (VTC) has not been widely used in polar processing studies. However, Mann et al. (2002), using a concentricity diagnostic different from that defined here, found that different values led to dramatically different patterns of denitrification in model simulations of NAT particle growth and evaporation because differences in the position of the cold region relative to the strong winds bounding the vortex can result in large differences in the amount of time air spends in cold regions of comparable size (Manney et al., 2003). Using idealized model simulations with a constant vortex field and cold regions ranging from highly concentric to nonconcentric, ~~Mann et al.~~ Mann et al. (2002) showed that concentricity affected denitrification independent of any variations in the vortex itself. Since ~~the concentricity calculation requires both the~~ our concentricity calculations require both the polar vortex and a cold region to be defined, intercomparisons of the resulting discontinuous time series can be difficult to interpret. Therefore, we present this diagnostic as occurrence frequencies of the range of VTC values, normalized by the total number of days on which VTC was calculated for each reanalysis. Figure ~~14~~ 16 shows the results at the 490 K level. The total number of days shown corresponds to the number of days with a valid VTC value summed over the years from the comparison time periods defined in Figure 1 and used in the monthly ~~relative bias~~ CPAD plots. As evidenced by the small regions of non-overlapping colors, in most cases the ERA-I vortices spend more time at higher concentricity values (within the 0.8 to 1.0 range) in both hemispheres. Consistent with the number of days below PSC thresholds shown in Figure ~~5~~6, ERA-I also tends to have more valid VTC days than MERRA in both hemispheres. The detailed implications of differences in concentricity are difficult to assess, but the results of Mann et al. (2002) and Manney et al. (2003) suggest that greater concentricity in ERA-I could ~~bias~~ lead model runs driven by that reanalysis ~~towards~~ to simulate longer-lasting PSCs, greater denitrification, and enhanced chlorine activation.

3.4 Trajectory Diagnostic Intercomparisons

Temperature histories along air parcel trajectories provide further information about polar processing potential beyond what can be obtained from the simple diagnostics described above. For these intercomparisons, we use isentropic trajectory calculations on the 490 K surface. As discussed by Manney et al. (2003), although neglecting cross-isentropic motion would not be suitable for detailed polar processing studies, this simplification allows us to efficiently run and analyze a large number of parcels for extensive intercomparisons. Following Manney et al. (2003, 2005a), we consider only the parcels that are initialized in regions with temperatures less than 195 K. With this initial filter, we examine the total and continuous amounts of time that these parcels spend at temperatures below 195 K (TT195 and CT195, respectively) before and after the initialization date, and the mean temperature of the parcels over the full run. The TT195 diagnostic captures the total exposure of an air mass to temperatures low enough for PSC formation and thus acts as a proxy for cumulative chlorine activation; CT195 better represents the potential lifetime of a PSC and is thus more directly relevant for denitrification, since sufficiently long continuous time at low temperatures allows PSC particles to grow to sizes large enough for sedimentation to occur (Manney et al., 2003, 2005a). Here we show trajectory calculations initialized on 10 January 1996 and 24 January 2011 for the NH, and 17 September 1988, 13 September 2002, and 25 May 2011 for the SH. These dates have been chosen to span a range of conditions representative of the interannual variability in both hemispheres, and in some cases to compare with previous studies using the same diagnostics. Several other cases have been investigated for each hemisphere; the results shown here are representative.

Figure ~~15~~ 17 summarizes the trajectory diagnostics for the NH. Overall, the parcel histograms for TT195 and CT195 are very similar between MERRA and ERA-I, with consistent distributions, peaks, and average values. The number of parcels included in the calculations indicate that MERRA has slightly larger regions of cold air, but the average values of the distributions only differ by approximately 0.05 days (∼1.2 hours) for TT195 and 0.5 days (∼12 hours) for CT195. Where there are small differences, they are consistent with minor spatial differences in maps of the parcels (not shown), which are also otherwise very similar. In investigating other NH winter initialization dates, we found that all agreed well (similar to the examples shown here) in each diagnostic for years as early as 1981, and as recent as 2013. The SH cases shown in Figure ~~16~~ 18 present a different picture. In particular, the mean parcel temperature plots reveal a significant improvement in agreement between MERRA and ERA-I over time. The 16 September 1988 case shows ~~a persistent warm bias in MERRA parcels~~ that MERRA parcels are consistently warmer, with temperatures roughly 1 K greater than those in ERA-I. The 13 September 2002 case also shows ~~a noticeable MERRA warm bias~~that MERRA parcels are noticeably warmer, but to a lesser extent than in 1988. Although we highlight a different part of the SH winter season for 2011, there is a lack of a discernible temperature ~~bias~~ differences in comparison to the two previous dates, ~~which agrees~~ in agreement with the results from Figures ~~3 and~~ 4 ~~; thus~~and 5, which showed that temperature differences between MERRA and

ERA-I are smallest from 2002-2013. Therefore, we consider the 25 May 2011 initialization date to be representative. The TT195 and CT195 histograms also reflect this evolution in the differences over the years, with the parcels on average spending fewer days below 195 K for MERRA than for ERA-I when the ~~warm bias in MERRA is present~~mean temperature of the MERRA parcels is higher.

The above results suggest two things. For the NH, isentropic trajectory runs initialized with MERRA or ERA-I should provide very similar results in the lower stratosphere, even for early years, thus lending confidence ~~in~~ that transport calculations using winds from these two reanalyses should give comparable outcomes. The same cannot be said for early years in the SH, where many fewer observations were available in the 1980s and 1990s, allowing differences in the models and DAS to dominate the data constraints. This resulted in systematic differences in temperature histories that are consistent with the large ~~relative biases~~ monthly CPADs seen in direct temperature diagnostics (e.g., Figures ~~3 through 7~~4 through 9). Overall, the agreement of the trajectory diagnostics from MERRA and ERA-I is much better than that from older analyses: In previous intercomparison studies, Manney et al. (2003, 2005a) found very large differences in trajectory runs driven by fields from earlier analyses/reanalyses. Discrepancies in 465 K trajectories for average TT195 were as large in magnitude as 5 days for the NH and 7 days for the SH, with discrepancies in average CT195 in both hemispheres up to 2.5 days. For 10 January 1996 (shown here in Figure ~~15~~17), the time series of average parcel histories calculated by Manney et al. (2003) showed differences as large as 10 K on some days; in addition, the five analyses compared in that study showed qualitatively very different distributions of TT195 and CT195. Large qualitative differences were also seen between the ~~different~~ analyses in September 2002 (Manney et al., 2005a), in contrast to the small ~~biases~~ differences and good qualitative agreement seen here in Figure ~~16~~ 18 (middle panels). The improvements in the agreement between MERRA and ERA-I over that between earlier analyses, which largely ingested the same data, demonstrate the degree of improvement in the models and DAS techniques over the past decade.

4 Discussion and Conclusions

We have presented comparisons of stratospheric polar processing diagnostics derived from the MERRA and ERA-Interim reanalyses for Arctic and Antarctic winters from 1979 through 2013. By using temperature, vortex, and trajectory diagnostics, we have comprehensively explored the major aspects of the dynamical fields that chemical destruction of polar ozone in the lower stratosphere is sensitive to. In addition, we have characterized how agreement between the two reanalyses evolved over the 1979-2013 period as assimilated observations changed. To do this, we compared the temperature and vortex diagnostics during five time periods bounded by large changes in the datasets that were assimilated. Most of the comparisons are shown using calculations of monthly ~~relative biases~~comparison period average differences (monthly CPADs), which are monthly means of the daily ~~average~~ dif-

ferences between MERRA and ERA-I averaged over the aforementioned time periods. Our primary conclusions are as follows:

- Comparisons of temperature diagnostics derived from MERRA and ERA-I show a major shift towards better agreement around 2002, especially at levels above about 490 K. At 580 K (around 30 hPa), ERA-I tends to have more days with lower temperatures, whereas MERRA tends to have larger regions of cold air. These results are consistent between the hemispheres.

- The comparisons of winter mean V_{PSC} have a complex dependence on time and altitude as evidenced by the A_{NAT} ~~relative biases~~monthly CPADs. The shift towards better agreement in both hemispheres for A_{NAT} above 490 K was not enough to make the comparisons of winter mean V_{PSC} agree better.

- Comparisons based on differences of winter mean V_{PSC}/V_{vort} can be complicated because of the dependence on vertical integrations and time averaging, and thus V_{PSC} and V_{vort} individually. Since MERRA tends to have larger regions of cold air (larger V_{PSC}) than ERA-I, and similarly or smaller-sized vortices in recent periods, MERRA also has larger vortex fractions of cold air in years beyond $\sim$1992.

- The vortex diagnostic comparisons are more complicated than those of the temperature diagnostics. In many cases the ~~relative biases do not improve~~monthly CPADs do not decrease in magnitude; some even increase over the 1979-2013 time period, especially in the Arctic. These ~~biases~~differences, however, tend to be small.

- Isentropic trajectory runs driven by MERRA and ERA-I give very similar results overall. We found that in the Northern Hemisphere, the trajectory diagnostics agree very well across most of the years, while in the Southern Hemisphere, the agreement improves significantly over time.

Overall, we found that agreement between MERRA and ERA-I is better in the Arctic than in the Antarctic for nearly all of the diagnostics, especially before approximately 2002. The monthly ~~relative biases~~ CPADs in the Southern Hemisphere show large differences in the first three periods, before the introduction of ATOVS data in 1998 (and subsequent introduction of Aqua data in 2002), but evolve over time to approach the level of agreement found in the ~~northern hemisphere~~Northern Hemisphere. Consistent behavior was also seen in our calculations of temperature histories from air parcel trajectories, in which agreement of mean parcel temperatures and distributions of the time spent below PSC formation thresholds improved substantially as more observations were introduced into the DAS. Nevertheless, even the relatively poor agreement between MERRA and ERA-I in the Southern Hemisphere during the earlier periods with sparse data is still considerably better than that between analyses and reanalyses available a decade ago. Furthermore, small ~~biases~~ differences are less critical to polar processing studies of the ~~SH than of the NH~~ Southern Hemisphere than those for

the Northern Hemisphere because the colder, more quiescent Antarctic winter conditions result in extensive (near total at some altitudes) ~~SH~~ Southern Hemisphere ozone destruction each ~~winter~~year.

The patterns and evolution of the differences between MERRA and ERA-I in the Arctic are much more complicated than those in the Antarctic. The temperature diagnostics in the NH show monthly ~~relative biases~~ CPADs decreasing in magnitude by a significant amount over the five observational periods studied, with, for example, maximum monthly mean differences in minimum temperature (the most sensitive diagnostic, as it relies on a single-point comparison for each day) since 1998 under 1 K, and no larger than ∼0.5 K after 2007. Since the development of PSCs depends critically on temperature thresholds, this close agreement means that choice of MERRA or ERA-I data is unlikely to make a substantial difference in most polar processing studies for time periods in the past 15 or so years. In contrast, the diagnostics of the strength and size of the NH polar vortex show ~~either~~ differences that either stay relatively constant or ~~even increasing biases~~ increase slightly over the years. This suggests that differences in the models and/or in the handling of the assimilated datasets can still be important factors even when the temperature fields are quite well constrained by data.

The results in this paper provide strong evidence that the agreement between MERRA and ERA-I evolves with their changing data inputs. While this is an unsurprising result, it confirms that changes in the assimilated observations often directly influence the analyzed temperature fields more than the model and assimilation characteristics do. Only when observations were sparse and nearly identical in MERRA and ERA-I (such as in the SH before 2002) did we see large differences that indicated the effect of model and assimilation system differences. Our results further indicate that ERA-I's assimilation of measurements from GPSRO and other additional instruments that are not used in MERRA in the final observation period (2007 through 2013) results in only a small improvement in stratospheric temperature diagnostics that already show good agreement after 2002. The most recent period has the best agreement for most of the diagnostics shown. This has been noted elsewhere: ~~Martineau and Son (2010a)~~ Martineau and Son (2010b) found that MERRA and ERA-I had the lowest biases among other reanalyses ~~in comparison~~ relative to COSMIC temperatures during the 2009 Arctic sudden stratospheric warming.

Further work is planned to more fully characterize the agreement of diagnostics of polar processing between recent reanalyses, and the importance of these diagnostics to polar processing studies. In the context of the SPARC (Stratosphere-troposphere Processes And their Role in Climate) Reanalysis Intercomparison Project (S-RIP; see http://s-rip.ees.hokudai.ac.jp/index.html), we plan to extend these intercomparisons to include the NCEP Climate Forecast System Reanalysis (NCEP/CFSR, Saha et al., 2010) and the Japanese 55-year Reanalysis ~~(JRA55, Ebita et al., 2011)~~ (JRA55, Kobayashi et al., 2015) ; these reanalyses, like MERRA and ERA-I, are recent high-resolution datasets that are valuable for numerous studies, including those of polar processing. In addition, two of the diagnostics introduced (sunlit vortex area and VTC) have not been widely used in previous polar processing studies. Work

in progress applying sunlit vortex area and VTC to disturbed versus more quiescent Arctic winters will help establish the sensitivity of polar processing and ozone loss to the conditions characterized by these diagnostics.

Acknowledgements. The authors would like to thank the personnel responsible for producing the ERA-Interim and MERRA reanalysis datasets. ERA-Interim data ~~was~~ were made available by ECMWF, Shinfield Park, Reading, UK; MERRA data ~~was~~ were provided by the GMAO, Greenbelt, Maryland, USA. Thanks to David Tan~~and Steven Pawsonfor~~ , Steven Pawson, and our reviewers, Simon Chabrillat and Craig Long, for their helpful comments and suggestions regarding this work~~,~~ ; Nathaniel Livesey for help with setting up the trajectory code~~,~~ ; and Brian Knosp, William Daffer, and the JPL Microwave Limb Sounder team for computing and data management support. Work at the Jet Propulsion Laboratory, California Institute of Technology, was done under contract with the National Aeronautics and Space Administration.

References

Berrisford, P., Dee, D. P., Fielding, K., Fuentes, M., Kallberg, P., Kobayashi, S., and Uppala, S.: The ERA-Interim Archive., ERA report series, pp. 1–16, 2009.

Bloom, S. C., Takacs, L. L., da Silva, A. M., and Ledvina, D.: Data Assimilation Using Incremental Analysis Updates, Mon. Weather Rev., 124, 1256–1271, 1996.

Brakebusch, M., Randall, C. E., Kinnison, D. E., Tilmes, S., Santee, M. L., and Manney, G. L.: Evaluation of whole atmosphere community climate model simulations of ozone during Arctic winter 2004–2005, J. Geophys. Res., 118, 2673–2688, 2013.

Charlton-Perez, A. J., Hawkins, E., Eyring, V., Cionni, I., Bodeker, G. E., Kinnison, D. E., Akiyoshi, H., Frith, S. M., Garcia, R., Gettelman, A., Lamarque, J. F., Nakamura, T., Pawson, S., Yamashita, Y., Bekki, S., Braesicke, P., Chipperfield, M. P., Dhomse, S., Marchand, M., Mancini, E., Morgenstern, O., Pitari, G., Plummer, D., Pyle, J. A., Rozanov, E., Scinocca, J., Shibata, K., Shepherd, T. G., Tian, W., and Waugh, D. W.: The potential to narrow uncertainty in projections of stratospheric ozone over the 21st century, Atmos. Chem. Phys., 10, 9473–9486, 2010.

Coy, L. and Pawson, S.: The Major Stratospheric Sudden Warming of January 2013: Analyses and Forecasts in the GEOS-5 Data Assimilation System, Mon. Weather Rev., 143, 491–510, doi:http://dx.doi.org/10.1175/MWR-D-14-00023.1, 2015.

Davies, S., Chipperfield, M. P., Carslaw, K. S., Sinnhuber, B.-M., Anderson, J. G., Stimpfle, R. M., Wilmouth, D. M., Fahey, D. W., Popp, P. J., Richard, E. C., von der Gathen, P., Jost, H., and Webster, C. R.: Modeling the effect of denitrification on Arctic ozone depletion during winter 1999/2000, J. Geophys. Res., 107, SOL–65–1Fe–SOL–65–18, 2002.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., and Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, Q. J. R. Meteorol. Soc., 137, 553–597, 2011.

Dunkerton, T. J. and Delisi, D. P.: Evolution of potential vorticity in the winter stratosphere of January-February 1979, J. Geophys. Res., 91, 1199–1208, 1986.

Ebita, A. et al.: The Japanese 55-year Reanalysis "JRA-55": an interim report, SOLA, 7, 149–152, 2011.

Feng, W., Chipperfield, M. P., Roscoe, H. K., Remedios, J. J., Waterfall, A. M., Stiller, G. P., Glatthor, N., Höpfner, M., and Wang, D.-Y.: Three-dimensional model study of the Antarctic ozone hole in 2002 and comparison with 2000, J. Atmos. Sci., 62, 822–837, 2005.

Feng, W., Chipperfield, M. P., Davies, S., von der Gathen, P., Kyrö, E., Volk, C. M., Ulanovsky, A., and Belyaev, G.: Large chemical ozone loss in 2004/2005 Arctic winter/spring, Geophysical Research Letters, 34, doi:10.1029/2006GL029098, 2007.

Fueglistaler, S., Liu, Y. S., Flannaghan, T. J., Haynes, P. H., Dee, D. P., Read, W. J., Remsberg, E. E., Thomason, L. W., Hurst, D. F., Lanzante, J. R., and Bernath, P. F.: The relation between atmospheric humidity and temperature trends for stratospheric water, J. Geophys. Res., 118, 1052–1074, 2013.

Fujiwara, M., Polavarapu, S., and Jackson, D.: A Proposal of the SPARC Reanalysis/Analysis Intercomparison Project, SPARC Newsletter, pp. 14 – 17, 2012.

Hanson, D. and Mauersberger, K.: Laboratory studies of the nitric acid trihydrate: Implications for the south polar stratosphere, Geophys. Res. Lett., 15, 855–858, 1988.

Harris, N. R., Lehmann, R., Rex, M., and von der Gathen, P.: A closer look at Arctic ozone loss and polar stratospheric clouds, Atmos. Chem. Phys., 10, 8499–8510, 2010.

Hitchcock, P., Shepherd, T., and Mclandress, C.: Past and future conditions for polar stratospheric cloud formation simulated by the Canadian Middle Atmosphere Model, Atmos. Chem. Phys., 9, 483–495, 2009.

Knox, J. A.: On converting potential temperature to altitude in the middle atmosphere, Eos Trans. AGU, 79, 376–378, 1998.

Kobayashi, S., Ota, Y., Harada, Y., Ebita, A., Moriya, M., Onoda, H., Onogi, K., Kamahori, H., Kobayashi, C., Endo, H., Miyaoka, K., and Takahashi, K.: The JRA-55 Reanalysis: General Specification and Basic Characteristics, J. Meteor. Soc. Japan, 93, doi:10.2151/jmsj.2015-001, 2015.

Livesey, N. J.: Aura Microwave Limb Sounder Lagrangian Trajectory Diagnostics Users' guide and file description document, Tech. rep., JPL, available from http://mls.jpl.nasa.gov/data/ltd.php, 2013.

Livesey, N. J., Santee, M. L., and Manney, G. L.: A Match-based approach to the estimation of polar stratospheric ozone loss using Aura Microwave Limb Sounder observations, Atmos. Chem. Phys. Disc., in press, 2015.

Lucchesi, R.: File Specification for MERRA products, GMAO Office Note No. 1 (Version 2.3), availablefromhttp://gmao.gsfc.nasa.gov/pubs/office_notes/, 2012.

Mann, G. W., Davies, S., Carslaw, K. S., Chipperfield, M. P., and Kettleborough, J.: Polar vortex concentricity as a controlling factor in Arctic denitrification, J. Geophys. Res., 107, 4663, doi:10.1029/2002JD002102, 2002.

Manney, G. L., Zurek, R. W., Gelman, M. E., Miller, A. J., and Nagatani, R.: The anomalous Arctic lower stratospheric polar vortex of 1992–1993, Geophys. Res. Lett., 21, 2405–2408, 1994a.

Manney, G. L., Zurek, R. W., O'Neill, A., and Swinbank, R.: On the motion of air through the stratospheric polar vortex, J. Atmos. Sci., 51, 2973–2994, 1994b.

Manney, G. L., Sabutis, J. L., Pawson, S., Santee, M. L., Naujokat, B., Swinbank, R., Gelman, M. E., and Ebisuzaki, W.: Lower stratospheric temperature differences between meteorological analyses in two cold Arctic winters and their impact on polar processing studies, J. Geophys. Res., 108, 8328, doi:10.1029/2001JD001149, 2003.

Manney, G. L., Allen, D. R., Krüger, K., Naujokat, B., Santee, M. L., Sabutis, J. L., Pawson, S., Swinbank, R., Randall, C. E., Simmons, A. J., and Long, C.: Diagnostic Comparison of Meteorological Analyses during the 2002 Antarctic Winter, Mon. Weather Rev., 133, 1261–1278, 2005a.

Manney, G. L., Krüger, K., Sabutis, J. L., Sena, S. A., and Pawson, S.: The remarkable 2003-2004 winter and other recent warm winters in the Arctic stratosphere since the late 1990s, J. Geophys. Res., 110, D04107, doi:10.1029/2004JD005367, 2005b.

Manney, G. L., Daffer, W. H., Zawodny, J. M., Bernath, P. F., Hoppel, K. W., Walker, K. A., Knosp, B. W., Boone, C., Remsberg, E. E., Santee, M. L., Harvey, V. L., Pawson, S., Jackson, D. R., Deaver, L., McElroy, C. T., McLinden, C. A., Drummond, J. R., Pumphrey, H. C., Lambert, A., Schwartz, M. J., Froidevaux,

L., McLeod, S., Takacs, L. L., Suarez, M. J., Trepte, C. R., Cuddy, D. C., Livesey, N. J., Harwood, R. S., and Waters, J. W.: Solar occultation satellite data and derived meteorological products: Sampling issues and comparisons with Aura Microwave Limb Sounder, J. Geophys. Res., 112, doi:10.1029/2007JD008709, 2007.

Manney, G. L., Santee, M. L., Rex, M., Livesey, N. J., Pitts, M. C., Veefkind, P., Nash, E. R., Wohltmann, I., Lehmann, R., Froidevaux, L., Poole, L. R., Schoeberl, M. R., Haffner, D. P., Davies, J., Dorokhov, V., Gernandt, H., Johnson, B., Kivi, R., Kyrö, E., Larsen, N., Levelt, P. F., Makshtas, A., McElroy, C. T., Nakajima, H., Parrondo, M. C., Tarasick, D. W., von der Gathen, P., Walker, K. A., and Zinoviev, N. S.: Unprecedented Arctic Ozone Loss in 2011, Nature, 478, 469–475, 2011.

Martineau, P. and Son, S.-W.: Quality of reanalysis data during stratospheric vortex weakening and intensification events, Geophys. Res. Lett., 37, doi:10.1029/2010GL045237, 2010a.

Martineau, P. and Son, S.-W.: Quality of reanalysis data during stratospheric vortex weakening and intensification events, Geophys. Res. Lett., 37, doi:10.1029/2010GL045237, 2010b.

McNally, A. P., Watts, P. D., A. Smith, J., Engelen, R., Kelly, G. A., Thépaut, J. N., and Matricardi, M.: The assimilation of AIRS radiance data at ECMWF, Q. J. R. Meteorol. Soc., 132, 935–957, doi:10.1256/qj.04.171, 2006.

Mitchell, D. M., Charlton-Perez, A. J., and Gray, L. J.: Characterizing the Variability and Extremes of the stratospheric polar vortices using 2D moment analysis, J. Atmos. Sci., 68, 1194–1213, 2011.

Pawson, S.: Representation of the Middle–to–Upper Stratosphere in MERRA, Global Modeling and Assimilation Office Annual Report & Research Highlights, p. 42, 2012.

Pawson, S. and Naujokat, B.: The cold winters of the middle 1990s in the northern lower stratosphere, J. Geophys. Res., 104, 14,209–14,222, 1999.

Pommereau, J.-P., Goutail, F., Lefèvre, F., Pazmino, A., Adams, C., Dorokhov, V., Eriksen, P., Kivi, R., Stebel, K., Zhao, X., and van Roozendael, M.: Why unprecedented ozone loss in the Arctic in 2011? Is it related to climate change, Atmos. Chem. Phys, 13, 5299–5308, 2013.

Rex, M., Von Der Gathen, P., Braathen, G., Harris, N., Reimer, E., Beck, A., Alfier, R., Krüger-carstensen, R., Chipperfield, M., De Backer, H., Balis, D., O'Connor, F., Dier, H., Dorokhov, V., Fast, H., Gamma, A., Gil, M., Kyrö, E., Litynska, Z., Mikkelsen, I., Molyneux, M., Murphy, G., Reid, S., Rummukainen, M., and Zerefos, C.: Chemical Ozone Loss in the Arctic Winter 1994/95 as Determined by the Match Technique, Journal of Atmospheric Chemistry, 32, 35–59, doi:10.1023/A:1006093826861, 1999.

Rex, M., Salawitch, R. J., Gathen, P., Harris, N. R., Chipperfield, M. P., and Naujokat, B.: Arctic ozone loss and climate change, Geophys. Res. Lett., 31, L04116, doi:10.1029/2003GL018844, 2004.

Rex, M., Salawitch, R. J., Deckelmann, H., von der Gathen, P., Harris, N. R. P., Chipperfield, M. P., Naujokat, B., Reimer, E., Allart, M., Andersen, S. B., Bevilacqua, R., Braathen, G. O., Claude, H., Davies, J., De Backer, H., Dier, H., Dorokhov, V., Fast, H., Gerding, M., Godin-Beekmann, S., Hoppel, K., Johnson, B., Kyrö, E., Litynska, Z., Moore, D., Nakane, H., Parrondo, M. C., Risley, A. D., Skrivankova, P., Stübi, R., Viatte, P., Yushkov, V., and Zerefos, C.: Arctic winter 2005: Implications for stratospheric ozone loss and climate change, Geophys. Res. Lett., 33, L23808, doi:10.1029/2006GL026731, 2006.

Rieder, H. and Polvani, L. M.: Are recent Arctic ozone losses caused by increasing greenhouse gases?, Geophys. Res. Lett., 40, 4437–4441, 2013.

Rienecker, M. 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, F. R., Ruddick, A. G., Sienkiewicz, M., and Woollen, J.: MERRA: NASA's modern-era retrospective analysis for research and applications, J. Clim., 24, 3624–3648, 2011.

Saha, S., Moorthi, S., Pan, H.-L., Wu, X., Wang, J., Nadiga, S., Tripp, P., Kistler, R., Woollen, J., Behringer, D., Liu, H., Stokes, D., Grumbine, R., Gayno, G., Wang, J., Hou, Y.-T., Chuang, H.-Y., Juang, H.-M. H., Sela, J., Iredell, M., Treadon, R., Kleist, D., Delst, P. V., Keyser, D., Derber, J., Ek, M., Meng, J., Wei, H., Yang, R., Lord, S., van den Dool, H., Kumar, A., Wang, W., Long, C., Chelliah, M., Xue, Y., Huang, B., Schemm, J.-K., Ebisuzaki, W., Lin, R., Xie, P., Chen, M., Zhou, S., Higgins, W., Zou, C.-Z., Liu, Q., Chen, Y., Han, Y., Cucurull, L., Reynolds, R. W., Rutledge, G., and Goldberg, M.: The NCEP Climate Forecast System Reanalysis, Bull. Am. Meteor. Soc., 91, 1015–1057, 2010.

Santee, M. L., Tabazadeh, A., Manney, G. L., Fromm, M. D., Jensen, E. J., Bevilacqua, R. M., and Waters, J. W.: A Lagrangian approach to studying Arctic polar stratospheric clouds using UARS MLS HNO_3 and POAM II aerosol extinction measurements, J. Geophys. Res., 107, 10.1029/2000J, doi:10.1029/2000JD000227, 2002.

Schoeberl, M. R. and Dessler, A. E.: Dehydration of the stratosphere, Atmos. Chem. Phys., 11, 8433–8446, 2011.

Schoeberl, M. R., Lait, L. R., Newman, P. A., and Rosenfield, J. E.: The structure of the polar vortex, J. Geophys. Res., 97, 7859–7882, 1992.

Schwartz, M. J., Lambert, A., Manney, G. L., Read, W. G., Livesey, N. J., Froidevaux, L., Ao, C. O., Bernath, P. F., Boone, C. D., Cofield, R. E., Daffer, W. H., Drouin, B. J., Fetzer, E. J., Fuller, R. A., Jarnot, R. F., Jiang, J. H., Jiang, Y. B., Knosp, B. W., Krüger, K., Li, J.-L. F., Mlynczak, M. G., Pawson, S., Russell, J. M., Santee, M. L., Snyder, W. V., Stek, P. C., Thurstans, R. P., Tompkins, A. M., Wagner, P. A., Walker, K. A., Waters, J. W., and Wu, D. L.: Validation of the Aura Microwave Limb SounderSic temperature and geopotential height measurements, J. Geophys. Res., 113, doi:10.1029/2007JD008783, http://dx.doi.org/10.1029/2007JD008783, 2008.

Seviour, W. J. M., Mitchell, D. M., and Gray, L. J.: A practical method to identify displaced and split stratospheric polar vortex events, Geophys. Res. Lett., 40, 5268–5273, doi:10.1002/grl.50927, 2013.

Sica, R. J., Izawa, M. R. M., Walker, K. A., Boone, C., Petelina, S. V., Argall, P. S., Bernath, P., Burns, G. B., Catoire, V., Collins, R. L., Daffer, W. H., De Clercq, C., Fan, Z. Y., Firanski, B. J., French, W. J. R., Gerard, P., Gerding, M., Granville, J., Innis, J. L., Keckhut, P., Kerzenmacher, T., Klekociuk, A. R., Kyrö, E., Lambert, J. C., Llewellyn, E. J., Manney, G. L., McDermid, I. S., Mizutani, K., Murayama, Y., Piccolo, C., Raspollini, P., Ridolfi, M., Robert, C., Steinbrecht, W., Strawbridge, K. B., Strong, K., Stübi, R., and Thurairajah, B.: Validation of the Atmospheric Chemistry Experiment (ACE) version 2.2 temperature using ground-based and space-borne measurements, Atmospheric Chemistry and Physics, 8, 35–62, doi:10.5194/acp-8-35-2008, http://www.atmos-chem-phys.net/8/35/2008/, 2008.

Simmons, A. J., Hortal, M., Kelly, G., McNally, A., Untch, A., and Uppala, S.: ECMWF analyses and forecasts of stratospheric winter polar vortex break-up: September 2002 in the southern hemisphere and related events, J. Atmos. Sci., 62, 668–689, 2005.

Simmons, A. J., Poli, P., Dee, D. P., Berrisfordand, P., Hersbach, H., Kobayashi, S., and Peubey, C.: Estimating low-frequency variability and trends in atmospheric temperature using ERA-Interim, Q. J. R. Meteorol. Soc., 140, 329–353, 2014.

Sinnhuber, B.-M., Stiller, G., Ruhnke, R., von Clarmann, T., Kellmann, S., and Aschmann, J.: Arctic winter 2010/2011 at the brink of an ozone hole, Geophys. Res. Lett., 38, doi:10.1029/2011GL049784, 2011.

Solomon, S.: Stratospheric ozone depletion: A review of concepts and history, Rev. Geophys., 37, 275–316, 1999.

Strahan, S. E., Douglass, A. R., and Newman, P. A.: The contributions of chemistry and transport to low Arctic ozone in March 2011 derived from Aura MLS observations, J. Geophys. Res., 118, 1563–1576, doi:10.1002/jgrd.50181, 2013.

Tilmes, S., Müller, R., Engel, A., Rex, M., and Russel, J. M.: Chemical ozone loss in the Arctic and Antarctic stratosphere between 1992 and 2005, Geophys. Res. Lett., 33, doi:10.1029/2006GL026925, 2006.

Waugh, D. W. and Randel, W. J.: Climatology of Arctic and Antarctic polar vortices using elliptical diagnostics, J. Atmos. Sci., 56, 1594–1613, 1999.

Wegner, T., Grooß, J.-U., von Hobe, M., Stroh, F., Sumińska-Ebersoldt, O., Volk, C. M., Hösen, E., Mitev, V., Shur, G., and Müller, R.: Heterogeneous chlorine activation on stratospheric aerosols and clouds in the Arctic polar vortex, Atmospheric Chemistry and Physics, 12, 11 095–11 106, doi:10.5194/acp-12-11095-2012, 2012.

WMO: Scientific assessment of ozone depletion: 2006, Global Ozone Res. and Monit. Proj. Rep. 50, Geneva, Switzerland, 2007.

WMO: Scientific assessment of ozone depletion: 2010, Global Ozone Res. and Monit. Proj. Rep. 52, Geneva, Switzerland, 2011.

WMO: Scientific assessment of ozone depletion: 2014, Global Ozone Res. and Monit. Proj. Rep. 55, Geneva, Switzerland, 2015.

Wohltmann, I. et al.: Uncertainties in modelling heterogeneous chemistry and Arctic ozone depletion in the winter 2009/2010, Atmos. Chem. Phys., 13, 3909–3929, 2013.

Table 1. Names of abbreviated instruments and satellites assimilated in ERA-Interim and MERRA.

Acronym	Full Name ~~Type~~
AIRS	Atmospheric Infrared Sounder ~~Instrument~~
AMSR	Advanced Microwave Scanning Radiometer ~~Instrument~~
AMSU	Advanced Microwave Sounding Unit ~~Instrument~~
ATOVS	Advanced Tiros Operational Vertical Sounder ~~System~~[1]
CHAMP	Challenging Minisatellite Payload ~~Satellite~~
COSMIC	Constellation Observing System for Meteorology, Ionosphere, and Climate ~~Satellite~~
GOES	Geostationary Operational Environmental Satellite ~~Satellites~~
HIRS	High resolution Infrared Radiation Sounder ~~Instrument~~
MHS	Microwave Humidity Sounder ~~Instrument~~
MSU	Microwave Sounding Unit ~~Instrument~~
SSM/I	Special Sensor Microwave Imager/Sounder ~~Instrument~~
SSU	Stratospheric Sounding Unit ~~Instrument~~
TOVS	Tiros Operational Vertical Sounder ~~System~~[1]

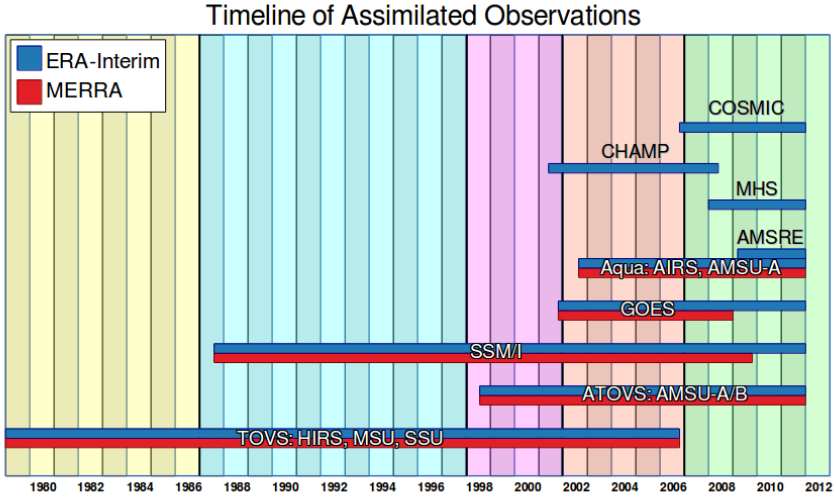

Figure 1. Timeline of satellite and GPSRO measurements assimilated in MERRA and ERA-Interim, organized by missions (when applicable) and instruments. Colored periods indicate regions of interest for intercomparisons, defined by significant changes in both data streams.

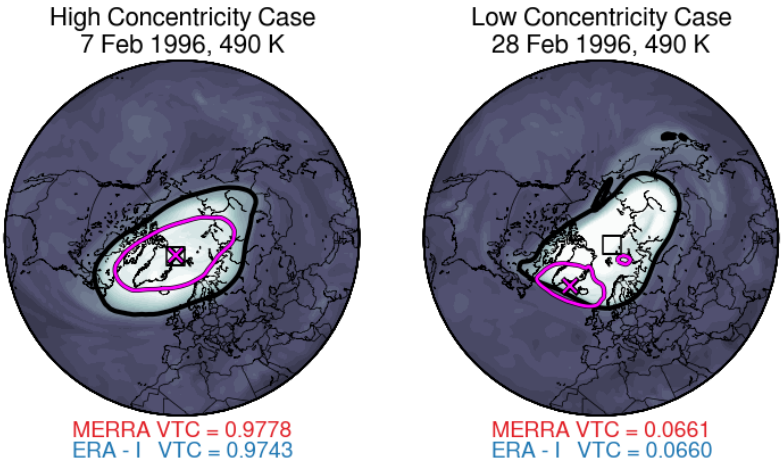

Figure 2. Maps showing cases when MERRA data indicate high (left) and low (right) values of vortex-temperature concentricity (VTC) at 490 K. The background field is potential vorticity from MERRA, and the thick black lines represent the vortex edge as defined by the 1.4×10^{-4} s^{-1} sPV contour. The magenta lines represent cold regions where temperatures (also from MERRA) are below T_{NAT}. The black squares and the magenta X marks are plotted at the centroid locations of the polar vortex and cold regions, respectively. For convenience, the corresponding values of VTC from MERRA (red) and ERA-Interim (blue) are shown below the maps.

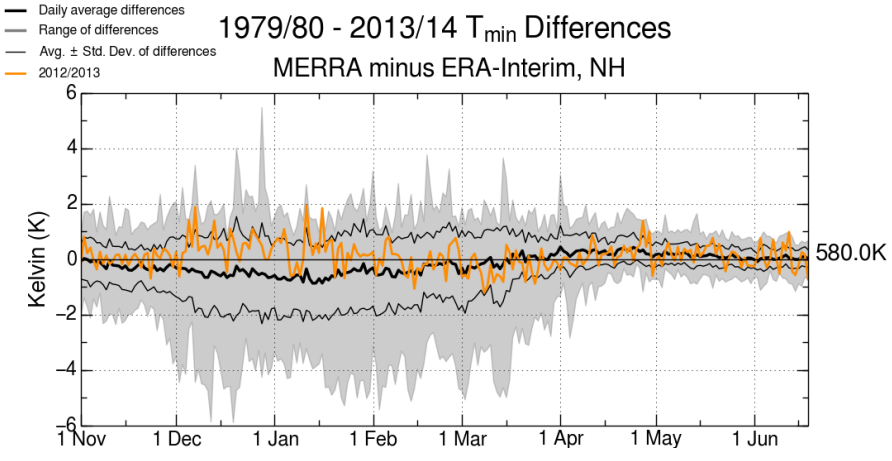

Figure 3. An example time series of the differences that are calculated for the daily diagnostics (T_{min}, A_{PSC}, etc.) that are in turn used to calculate the monthly ~~relative biases~~comparison period average differences (monthly CPADs). This case shows the differences in T_{min} between MERRA and ERA-I at 580 K for the 2012/2013 Arctic winter in orange, with the range of differences over all NH winters from 1979-2013 shown by the grey envelope. The thick black line represents the average daily differences over all years, while the thin black lines show one standard deviation. Since the average daily differences vary over the ~~seasons~~ season shown, we compute the monthly ~~relative biases~~ CPADs as the monthly means of ~~the~~ these average daily differences.

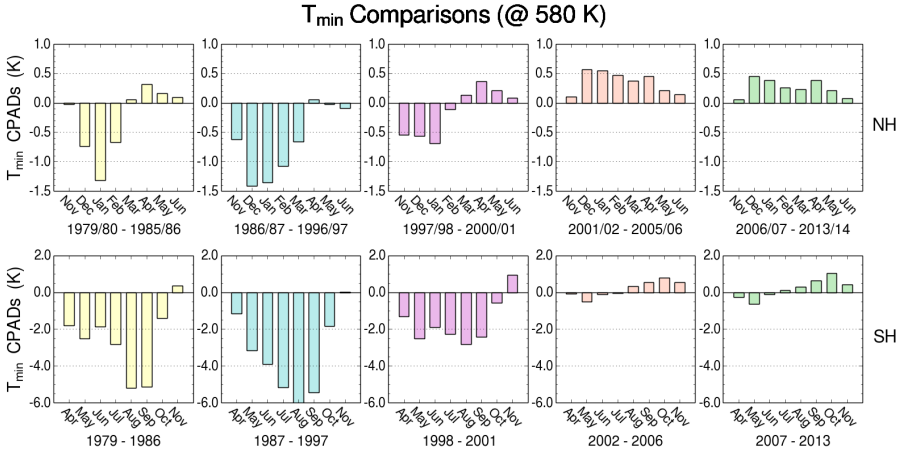

Figure 4. Monthly ~~relative biases in~~ CPADs of T_{min} between MERRA and ERA-Interim at 580 K potential temperature for Arctic (top) and Antarctic (bottom) winters. The colors used for the columns correspond to the colored regions of Figure 1. The NH (SH) panels cover months from November through June (April through November). Note that the T_{min} (y-axis) scales are different for each hemisphere.

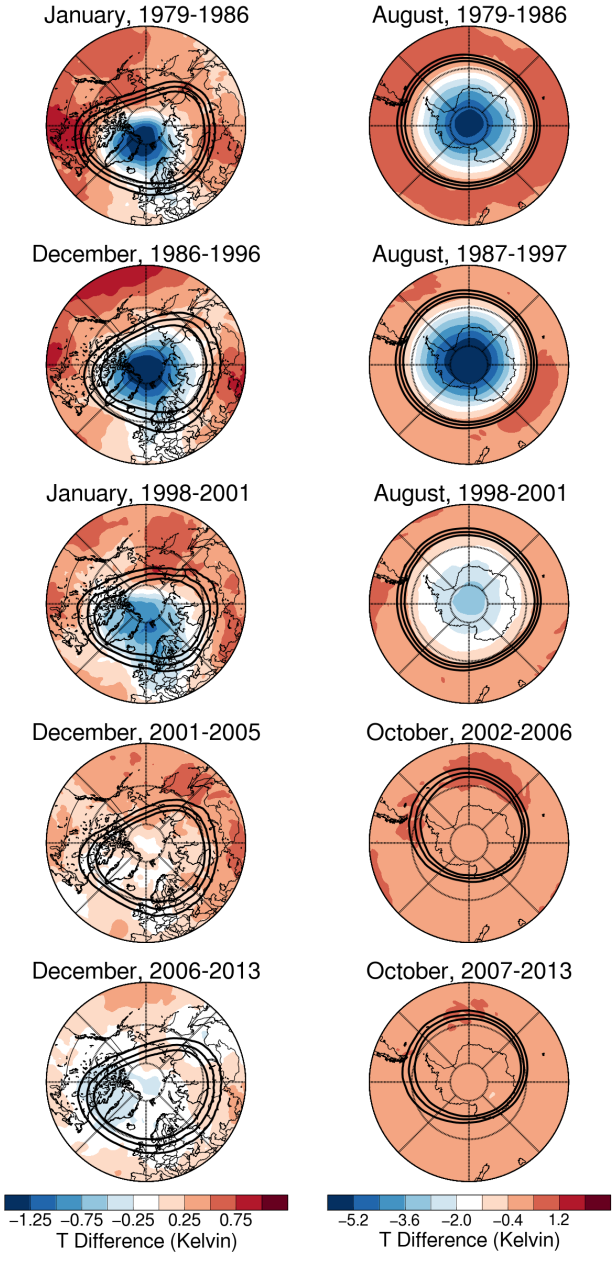

Figure 5. Maps of mean MERRA minus ~~ERA-Interim~~ ERA-I 12UT temperatures (averaged over the years listed) at 580 K potential temperature for the Arctic (left) and Antarctic (right) poleward of 40°. Contours of scaled potential vorticity (1.4, 1.6, and 1.8 $\times$ 10^{-4} s^{-1}) from ERA-Interim in the vortex edge region are overlaid in black. The months shown correspond to ~~the months~~ those with the largest magnitude T_{min} ~~relative biases, and the~~ monthly CPADs. The temperature scales are the same as shown in Figure ~~3~~ 4 (which differ for the Arctic and Antarctic).

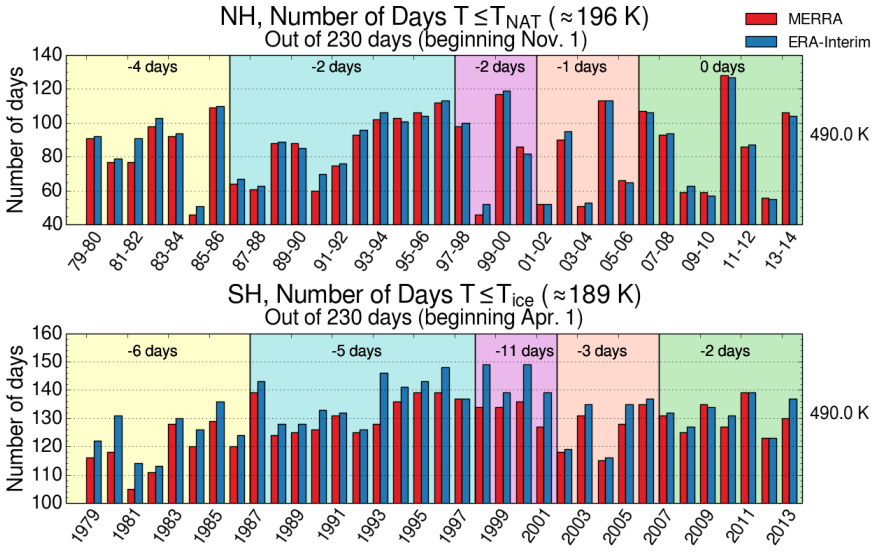

Figure 6. Number of days with Arctic (top) and Antarctic (bottom) winter minimum temperatures below NAT and ice PSC thresholds at 490 K. Note that the y-axes do not start from zero, and have different ranges for each hemisphere. The black numbers at the top of each colored region indicate the average differences (MERRA minus ERA-I) for the time-period, rounded to the nearest day.

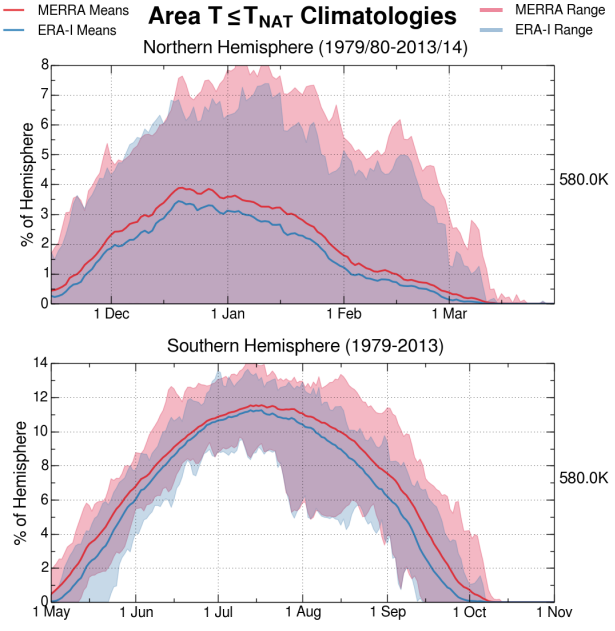

Figure 7. Mean values for A_{NAT} (thick red/blue lines), expressed as a percentage of a hemisphere, for Arctic and Antarctic winters at 580 K potential temperature. The blue (red) envelope shows the range of ERA-I (MERRA) values, with purple indicating where the ranges of the two reanalyses overlap.

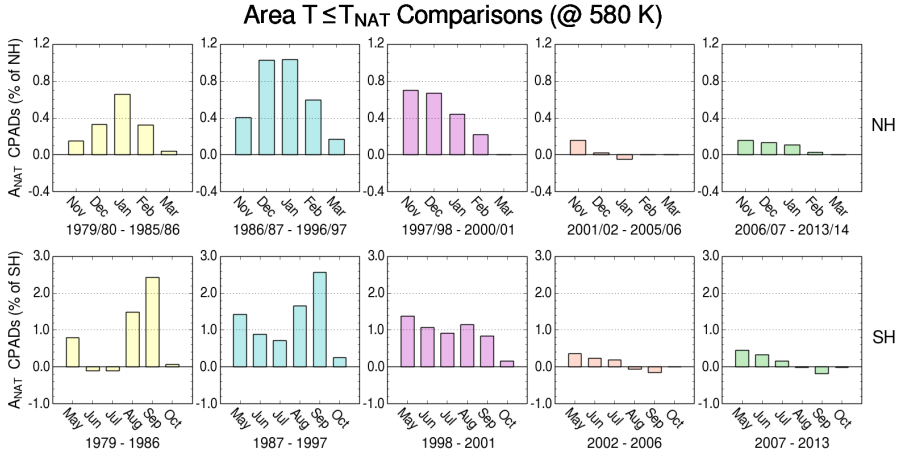

Figure 8. Monthly ~~relative biases in~~ CPADs of A_{NAT} (in % of a hemisphere) between MERRA and ERA-Interim at 580 K potential temperature for Arctic (top) and Antarctic (bottom) winters. Time periods are as in Figure ~~3.~~ 4. The NH (SH) panels cover months from November through March (May through October). The A_{NAT} (y-axis) scales are different for each hemisphere.

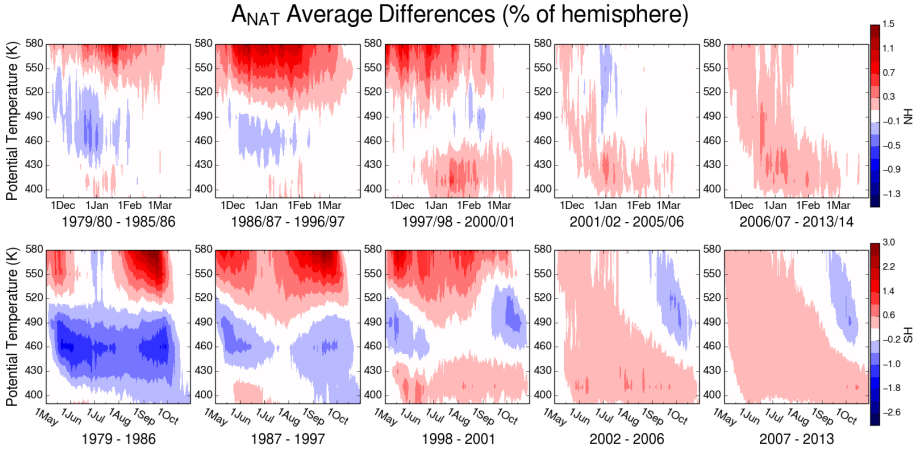

Figure 9. Differences in A_{NAT} (in % of a hemisphere) between MERRA and ERA-I from 390 to 580 K, averaged over the comparison periods of Figure 1 for Arctic (top row) and Antarctic (bottom row) winters.

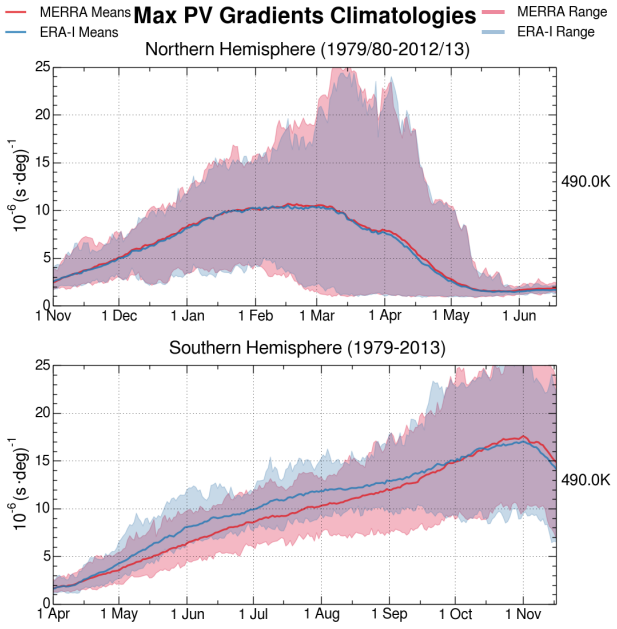

Figure 10. Mean values for ~~the area of the sunlit polar vortex~~ maximum PV gradients (thick red/blue lines) for Arctic and Antarctic winters at 490 K potential temperature. The blue (red) envelope shows the range of ERA-I (MERRA) values, with purple indicating where the ranges of the two reanalyses overlap.

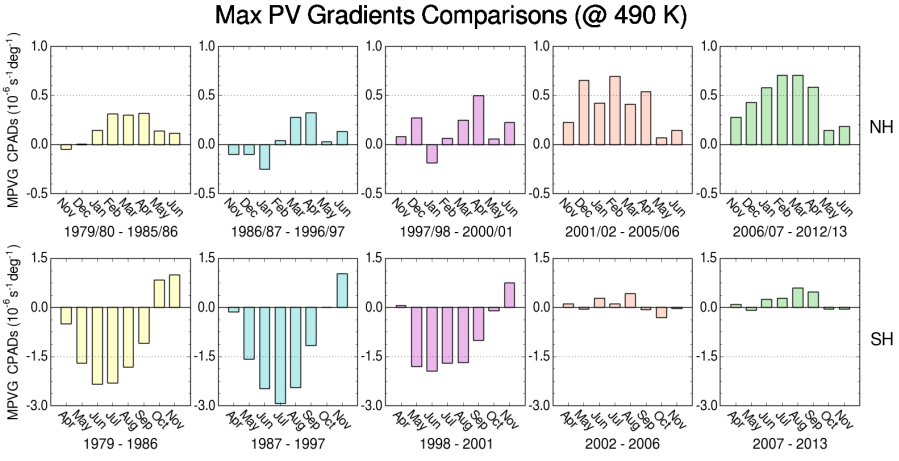

Figure 11. Monthly ~~relative biases in sunlit vortex area~~ CPADs of maximum potential vorticity gradients between MERRA and ERA-Interim at 490 K potential temperature for Arctic (top) and Antarctic (bottom) winters. Time periods are as in Figure ~~3.~~ 4. The NH (SH) panels cover months from November through ~~March~~ June (~~May~~ April through ~~October~~November). ~~Sunlit area~~ The MPVG (y-axis) scales are different for each hemisphere.

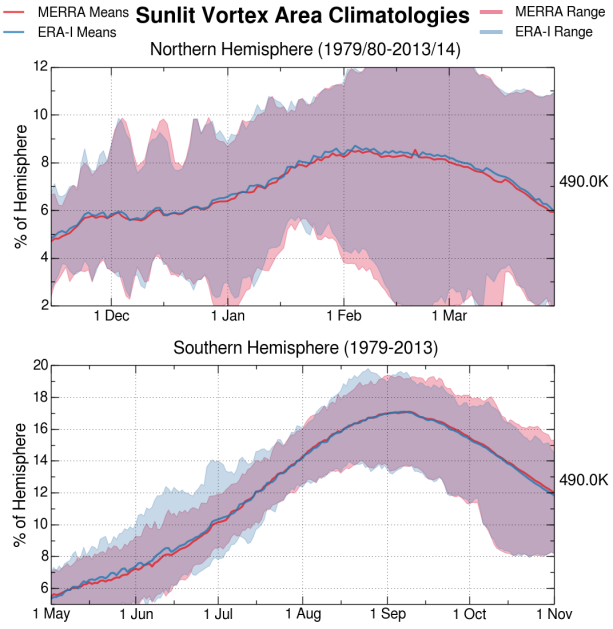

Figure 12. Mean values for the sunlit area of the polar vortex (thick red/blue lines), expressed as percentages of a hemisphere, for Arctic and Antarctic winters at 490 K potential temperature. The blue (red) envelope shows the range of ERA-I (MERRA) values, with purple indicating where the ranges of the two reanalyses overlap.

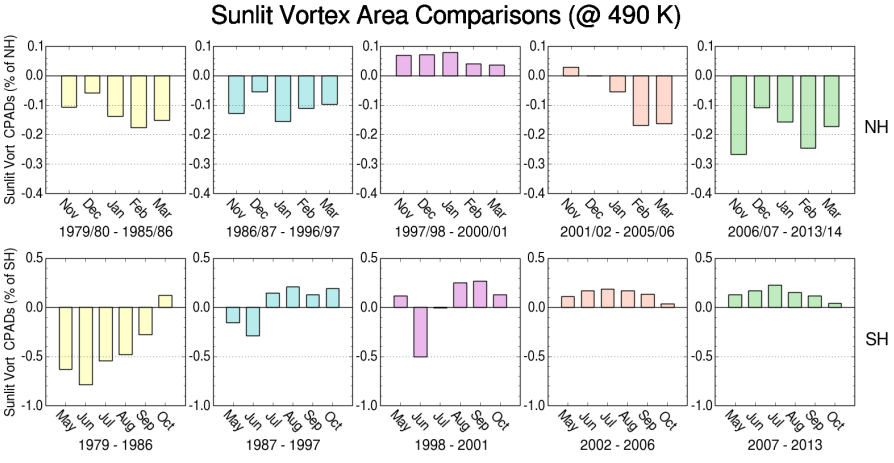

Figure 13. Monthly CPADs in sunlit vortex area between MERRA and ERA-Interim at 490 K potential temperature for Arctic (top) and Antarctic (bottom) winters. Time periods are as in Figure 4. The NH (SH) panels cover months from November through March (May through October). The sunlit area (y-axis) scales are different for each hemisphere.

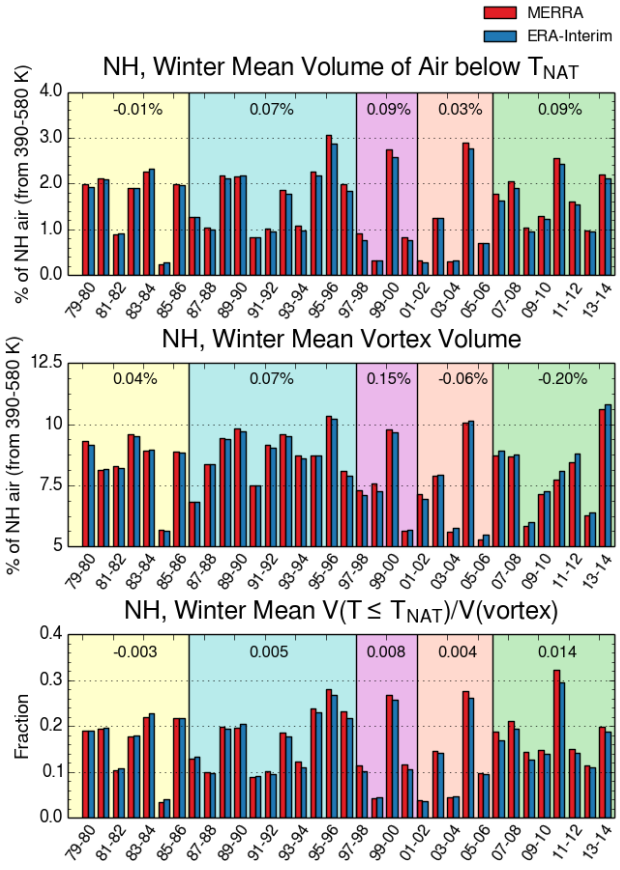

Figure 14. Arctic winter mean V_{PSC} (top), V_{vort} (center), and V_{PSC} expressed as a fraction of vortex volume (bottom). Note that the y-axis for winter mean V_{vort} does not start from zero. The black numbers at the top of each colored region indicate the average differences (MERRA minus ERA-I) for the time-period.

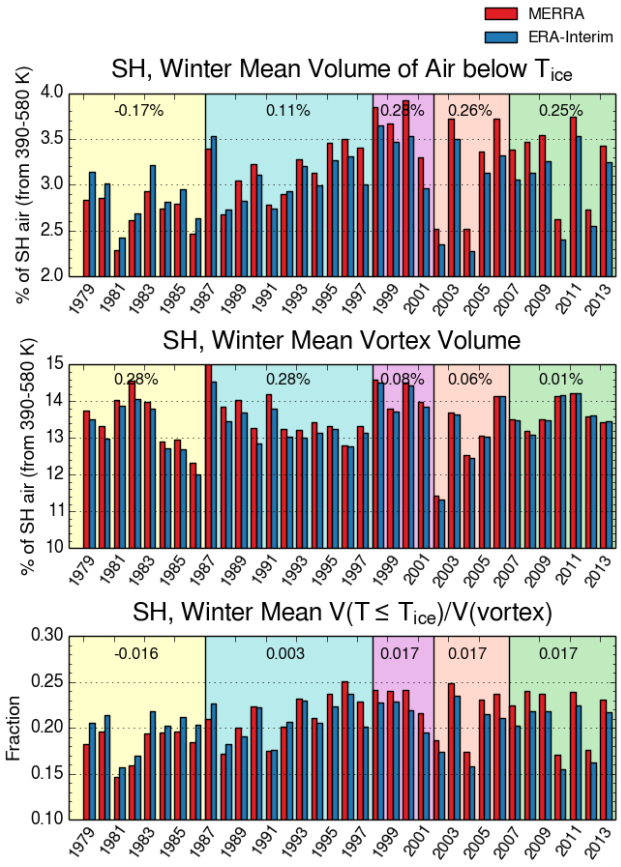

Figure 15. As in Figure ~~12~~14, but for the Antarctic. Also note that, here, none of the y-axes start from zero.

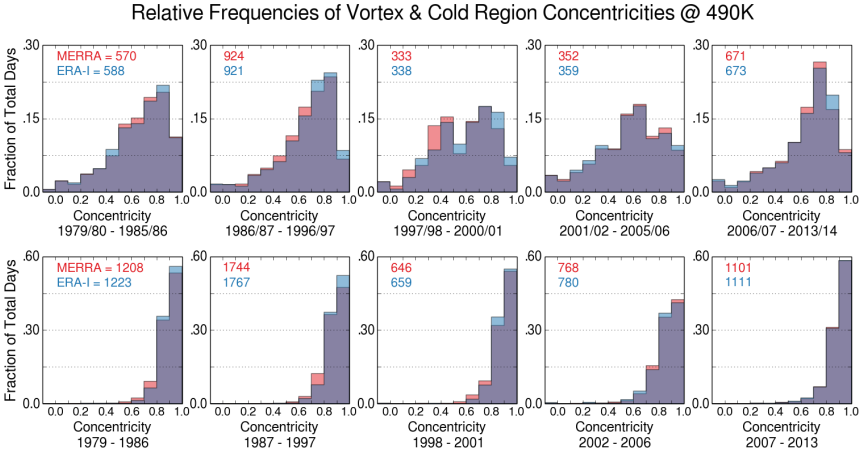

Figure 16. Concentricity of the polar vortex and regions with $T \leq T_{NAT}$ expressed as relative frequencies of the total number of days with a valid concentricity value (red and blue numbers) from the time periods of Figure 1 for Arctic (top) and Antarctic (bottom) winters.

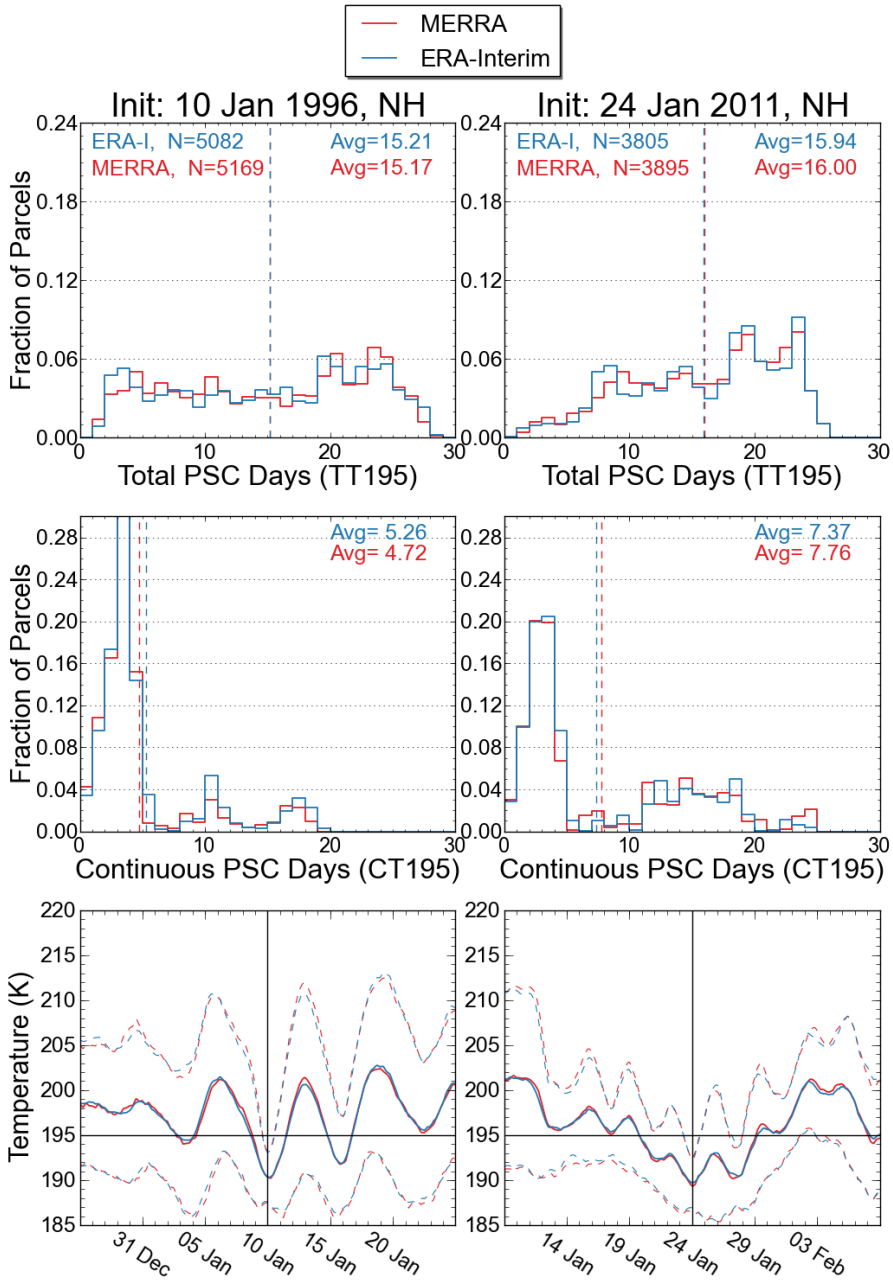

Figure 17. Temperature histories along air parcel trajectories at 490 K (∼56 hPa). Both of the columns are from Arctic winters. The first and second rows are histograms of parcels that spent total and continuous time in temperatures below 195 K (see text). The dashed vertical lines show the averages of TT195 and CT195 for each of the reanalyses. The last row is the mean temperature of the parcels initialized in the cold regions for the full 30 day (15 day forward/backward) trajectory runs; the black vertical line indicates the initialization date, while the black horizontal line marks 195 K. The dashed red and blue curves show one standard deviation range.

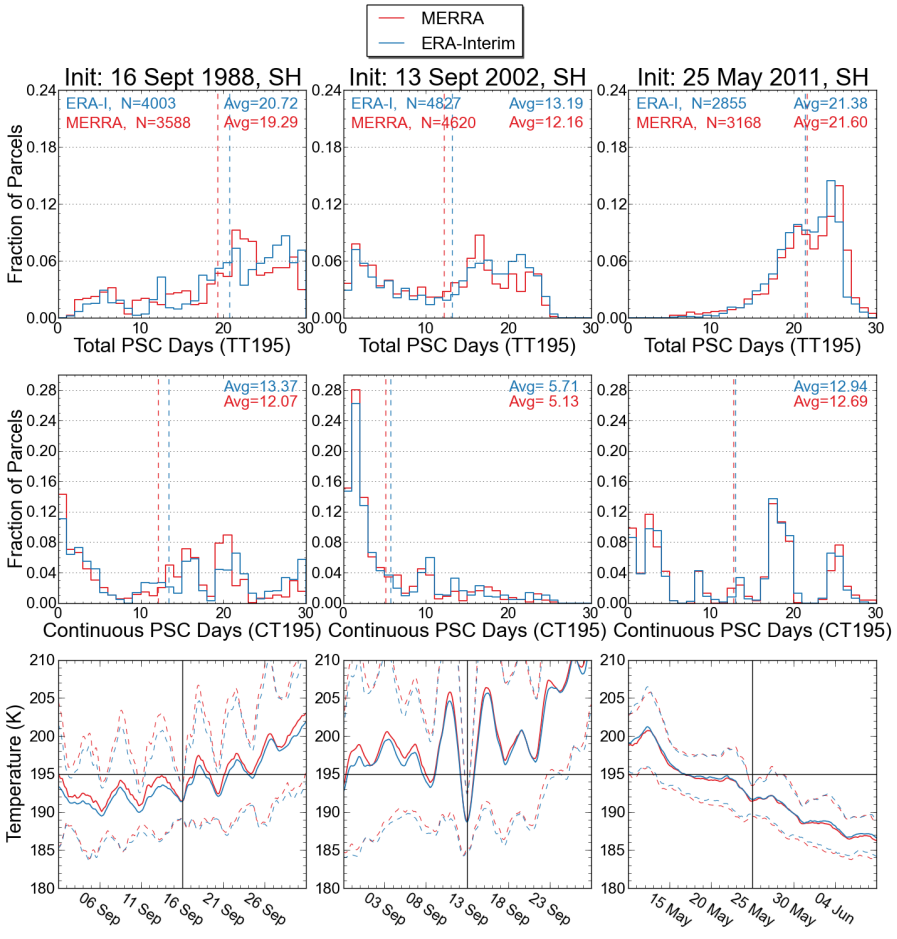

Figure 18. As in Figure ~~15~~17, but for Antarctic winters.