Chapter 12: Synthesis Summary

Chapter authors

Masatomo Fujiwara	Faculty of Environmental Earth Science, Hokkaido University	Japan
Gloria L. Manney	(1) NorthWest Research Associates (2) Department of Physics, New Mexico Institute of Mining and Technology	USA
Lesley J. Gray	(1) Atmospheric, Oceanic and Planetary Physics, University of Oxford (2) NERC National Centre for Atmospheric Science	UK
Jonathon S. Wright	Department of Earth System Science, Tsinghua University	China
James Anstey	Canadian Centre for Climate Modelling and Analysis Environment and Climate Change Canada	Canada
Thomas Birner	(1) Meteorological Institute, Ludwig-Maximilians-University Munich (2) Institute of Atmospheric Physics, German Aerospace Center (DLR Oberpfaffenhofen) *previously at: Department of Atmospheric Science, Colorado State University, USA*	Germany
Sean Davis	Chemical Sciences Laboratory, National Oceanic and Atmospheric Administration	USA
Edwin P. Gerber	Courant Institute of Mathematical Sciences, New York University	USA
V. Lynn Harvey	Laboratory for Atmospheric and Space Physics, University of Colorado	USA
Michaela I. Hegglin	Department of Meteorology, University of Reading	UK
Cameron R. Homeyer	School of Meteorology, University of Oklahoma	USA
John A. Knox	Department of Geography, University of Georgia	USA
Kirstin Krüger	Department of Geosciences, University of Oslo	Norway
Alyn Lambert	Jet Propulsion Laboratory, California Institute of Technology	USA
Craig S. Long	Climate Prediction Center, National Oceanic and Atmospheric Administration *(retired)*	USA
Patrick Martineau	Japan Agency for Marine-Earth Science and Technology *previously at: Research Center for Advanced Science and Technology, the University of Tokyo*	Japan
Beatriz M. Monge-Sanz	Atmospheric, Oceanic and Planetary Physics, University of Oxford	UK
Michelle L. Santee	Jet Propulsion Laboratory, California Institute of Technology	USA
Susann Tegtmeier	Institute of Space and Atmospheric Studies, University of Saskatchewan *previously at: GEOMAR Helmholtz Centre for Ocean Research Kiel, Germany*	Canada
Simon Chabrillat	Royal Belgian Institute for Space Aeronomy	Belgium
Zachary D. Lawrence	(1) Cooperative Institute for Research in Environmental Sciences, Univ. of Colorado (2) NOAA Physical Sciences Laboratory (PSL) (3) NorthWest Research Associates	USA
David G. H. Tan	European Centre for Medium-Range Weather Forecasts *(retired)*	UK
David R. Jackson	Met Office	UK
Saroja Polavarapu	Climate Research Division, Environment and Climate Change Canada	Canada
Gilbert P. Compo	(1) Cooperative Institute for Research in Environmental Sciences, University of Colorado; (2) Physical Sciences Laboratory, National Oceanic and Atmospheric Administration	USA
Rossana Dragani	European Centre for Medium-Range Weather Forecasts	UK
Wesley Ebisuzaki	National Oceanic and Atmospheric Administration	USA
Yayoi Harada	Japan Meteorological Agency	Japan
Chiaki Kobayashi	Japan Meteorological Agency	Japan
Krzysztof Wargan	(1) National Aeronautics and Space Administration (2) Science Systems and Applications, Inc.	USA
Jeffrey S. Whitaker	National Oceanic and Atmospheric Administration	USA

Contents

12.1 Key findings and recommendations by chapter

This section lists the key findings and recommendations from each of *Chapters 3 - 11*. The key findings collectively provide a concise overview of the content and results for the corresponding chapter, while the recommendations include both guidelines for reanalysis data users and suggestions for reanalysis data producers. Each subsection also includes a summary figure assessing the reliability of selected reanalyses with respect to key diagnostics, which is reproduced from the summary section of the corresponding chapter. (Please refer to the footnote [1] for the meaning of the evaluation terms, *i.e.*, demonstrated suitable, suitable with limitations, use with caution, demonstrated unsuitable, and unevaluated.) These assessments, while inherently subjective, are intended to provide the reader with an overview of the relative quality of the diagnostic. So, for example, across a given diagnostic the relative performance of the different reanalyses can be compared, and (*e.g.*, for a given reanalysis) the performance across different diagnostics can be compared. Only those diagnostics specifically examined either in this report or in previously published papers are assigned a score in the table; otherwise they are marked unevaluated. Although not all diagnostics we use can be evaluated against observations, we attempt to assign evaluation scores to any key diagnostics that can be readily summarized. For those that cannot be compared to observations, our assessment reflects consistency with other processes and current understanding of the phenomenon in question. Diagnostics that preclude simple classification (*e.g.*, the assessment of polar transport processes yielded results that varied by hemisphere, time of year, altitude, location in the polar vortex, and species) are omitted from these summary figures. Readers interested in further information about any diagnostic should refer to the corresponding chapter and section of the report, for which a key is provided in each summary figure.

It is noted here that, as explained in *Chapter 1* (**Table 1.1**), some ERA5 data have been available since July 2018, ERA5 data from 1979 onward have been available since January 2019, and a preliminary version of ERA5 1950 - 1978 data have been available since November 2020. Because most of the studies in this report were finalized before ERA5 was readily available, full evaluation of ERA5 has not been made. However, Chapter 2 includes information on the ERA5 system, and some chapters show ERA5 results for some diagnostics.

A key for all abbreviations used in this chapter is provided at the end.

12.1.1 Chapter 3: Overview of Temperature and Winds

In this chapter, we have examined reanalysis representations of key diagnostics related to temperature and winds. A summary of the diagnostics evaluated in this chapter is provided in **Figure 12.1**, which also directs the reader to the appropriate chapter section for further information. Below, we briefly summarize the key findings from this chapter and recommendations for both the appropriate use of and potential for improving reanalysis temperature and wind fields.

Key Findings of Chapter 3:

- More recent reanalyses from all centres consistently outperform earlier versions. (*e.g.*, JRA-55 *vs*. JRA-25; MERRA-2 *vs*. MERRA).
- Drifts and jumps in the long-term temperature time series can occur due to changes in available data sources. These irregularities are most pronounced at altitudes above 10 hPa. Greatest caution is advised when determining trends with reanalysis temperature data sets above 10 hPa.
- The more recent reanalyses have fewer discontinuities in their temperature and wind time series owing to improved data assimilation techniques and smoother transitions among different sets of observations.

---

[1] (As in *Chapter 1*, *Section 1.3*)
Demonstrated suitable: the reanalysis product could be directly validated using observational or physical constraints and was found to be in close agreement with expectations
Suitable with limitations: the reanalysis product could be directly validated using observatial or physical constraints and exhibited limited agreement; or, appropriate constraints were unavailable but reanalysis products were consistent beyond specific limitations as described in the text
Use with caution: the reanalysis system contains all elements necessary to provide a useful representation of this variable or process, but that representation has evident red flags (*e.g.*, disagreement with available observations; meaningful disagreements among reanalyses that cannot be resolved at this point)
Demonstrated unsuitable: the reanalysis product has been flagged as unable to represent processes that are key for this diagnostic as assessed in this report or by previous studies. This category is reserved for situations where the reanalysis is missing something fundamental in its structure (*e.g.*, a model top at 3 hPa means NCEP-NCAR R1 is 'demonstrated unsuitable' for studying processes in the USLM)
Unevaluated: the performance of the reanalysis product with respect to this diagnostic or variable has not been examined in this report or by previous studies

- The transition from the TOVS to ATOVS satellite periods starting around 1998 - 1999 is problematic for all reanalyses. In the stratosphere, the transition from three broad SSU infrared channels to five narrower AMSU/ATMS microwave channels proves to be problematic for data assimilation.

- The more recent reanalyses agree quite well with each other in the lower and middle stratosphere. All reanalyses have greater differences in the upper stratosphere and lower mesosphere. The latter discrepancies result from differences in model top, vertical resolution, data assimilation techniques, and data that are assimilated. *Chapter 2* provides detailed information about each reanalysis system.

- Temperature biases exist between the various reanalyses in the UTLS, especially before 1998. Temperatures in this region do not harmonize until after 2005, when widespread GNSS-RO observations became available.

- The agreement between Singapore radiosonde winds and reanalysis QBO winds at Singapore is better in the second half than the first half of the 1980 - 2014 record, consistent with improved constraints on reanalysis winds due to the gradual increase in the number of radiosonde observations over time. We expect that future reanalyses will have better QBO winds as forecast models become better able to produce a spontaneous QBO in the tropics.

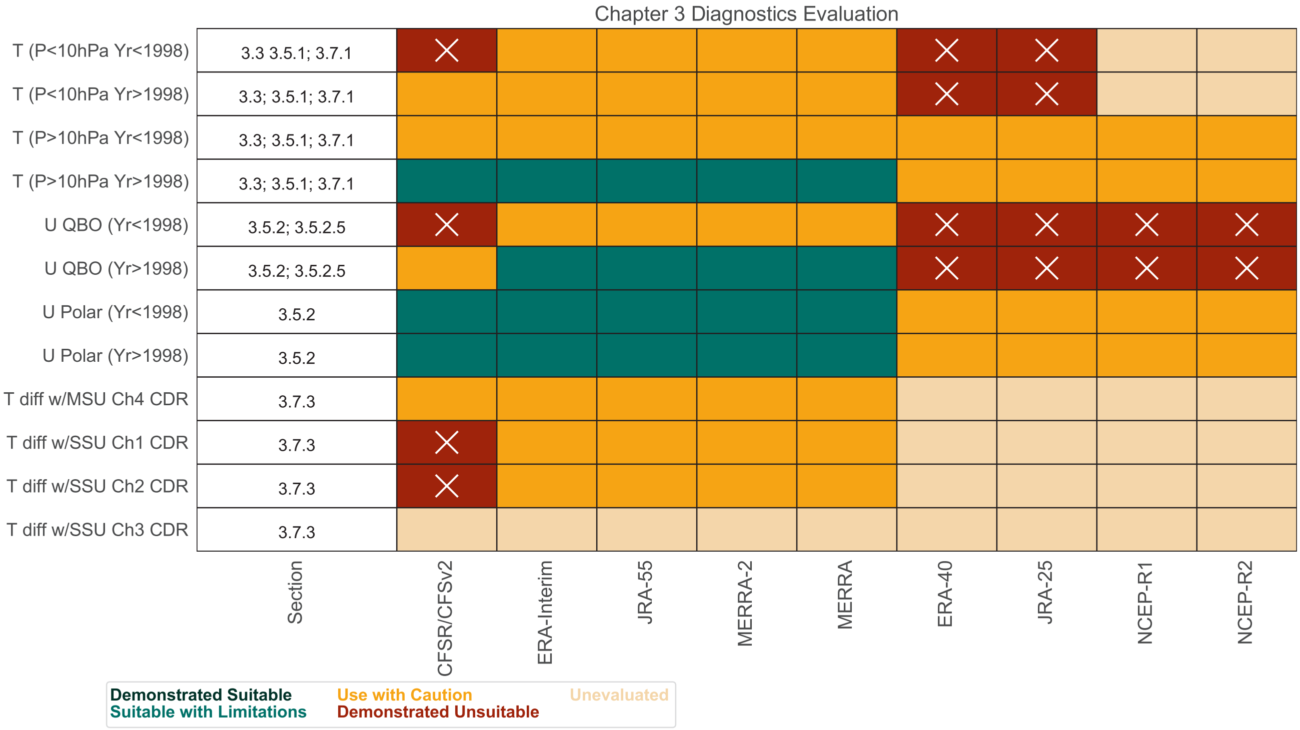
Recommendations from Chapter 3:

- Users of any reanalysis should proceed with greatest caution when intercomparing reanalyses, and particularly when attempting to detect trends and/or changes in climate above the tropopause (see also *Section 12.2*).

- Improving the TOVS period would be highly beneficial to future reanalyses, especially for climate studies. However, the TOVS period may never be as good as the ATOVS period due to the relative sparsity and coarser vertical resolution of the assimilated data.

- Improvements to the variational bias correction schemes for handling the broad SSU weighting functions and improvements to the forecast models (especially the non-orographic gravity wave parameterizations, so that forecast models can generate a realistic QBO on their own) are some of the ways the TOVS time period can be improved upon.

- It may benefit each "satellite-era" reanalysis to begin their reanalysis several years earlier using just conventional data. This most likely will help harmonize the reanalyses' temperature structure below 10 hPa at the start of assimilating satellite data.



Chapter 3 Diagnostics Evaluation

	Section	CFSR/CFSv2	ERA-Interim	JRA-55	MERRA-2	MERRA	ERA-40	JRA-25	NCEP-R1	NCEP-R2
T (P<10hPa Yr<1998)	3.3 3.5.1; 3.7.1	🟥✕	🟧	🟧	🟧	🟧	🟥✕	🟥✕	⬜	⬜
T (P<10hPa Yr>1998)	3.3; 3.5.1; 3.7.1	🟧	🟧	🟧	🟧	🟧	🟥✕	🟥✕	⬜	⬜
T (P>10hPa Yr<1998)	3.3; 3.5.1; 3.7.1	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧
T (P>10hPa Yr>1998)	3.3; 3.5.1; 3.7.1	🟩	🟩	🟩	🟩	🟩	🟧	🟧	🟧	🟧
U QBO (Yr<1998)	3.5.2; 3.5.2.5	🟥✕	🟧	🟧	🟧	🟧	🟥✕	🟥✕	🟥✕	🟥✕
U QBO (Yr>1998)	3.5.2; 3.5.2.5	🟧	🟩	🟩	🟩	🟩	🟥✕	🟥✕	🟥✕	🟥✕
U Polar (Yr<1998)	3.5.2	🟩	🟩	🟩	🟩	🟩	🟧	🟧	🟧	🟧
U Polar (Yr>1998)	3.5.2	🟩	🟩	🟩	🟩	🟩	🟧	🟧	🟧	🟧
T diff w/MSU Ch4 CDR	3.7.3	🟧	🟧	🟧	🟧	🟧	⬜	⬜	⬜	⬜
T diff w/SSU Ch1 CDR	3.7.3	🟥✕	🟧	🟧	🟧	🟧	⬜	⬜	⬜	⬜
T diff w/SSU Ch2 CDR	3.7.3	🟥✕	🟧	🟧	🟧	🟧	⬜	⬜	⬜	⬜
T diff w/SSU Ch3 CDR	3.7.3	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜

Demonstrated Suitable
Suitable with Limitations
Use with Caution
Demonstrated Unsuitable
Unevaluated

Figure 12.1: *(Same as **Figure 3.26**.) A summary of the diagnostics evaluated in Chapter 3: Overview of Temperature and Winds. The "Section" column at left indicates where in the chapter each diagnostic is described. See the beginning of this chapter for the meaning of the evaluation terms. Note that the score corresponding to "demonstrated suitable" was not assigned to any of the diagnostics listed here, so the darkest green colour does not appear in this table. "T" = temperature; "P" = pressure; "Yr" = year, "U" = zonal wind; "QBO" = Quasi-Biennial Oscillation; "T diff w/" = temperature difference with; "MSU" = Microwave Sounding Unit (a satellite instrument); "Ch" = Channel; "CDR" = climate data record; "SSU" = Stratospheric Sounding Unit (a satellite instrument).*

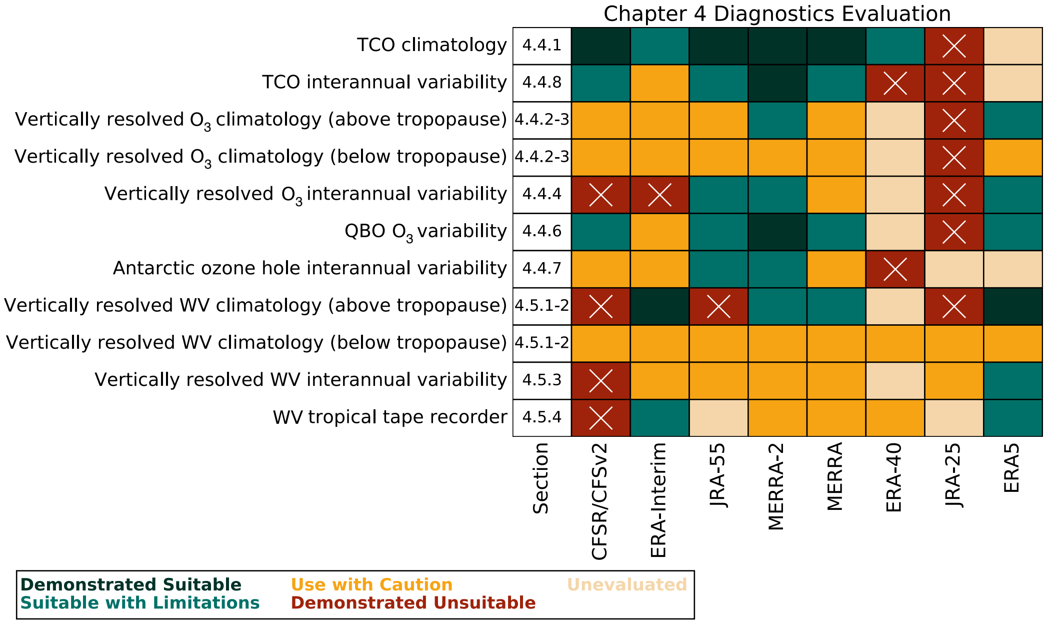


Figure 12.2: *(Same as **Figure 4.21**.) A summary of the diagnostics evaluated in Chapter 4: Overview of Ozone and Water Vapour. The "Section" column at left indicates where in the chapter each diagnostic is described. See the beginning of this chapter for the meaning of the evaluation terms. "TCO" = Total Column Ozone; "QBO" = Quasi-Biennial Oscillation; "WV" = water vapour.*

12.1.2 Chapter 4: Overview of Ozone and Water Vapour

In this chapter, we have assessed the reanalysis representations of key diagnostics related to ozone and water vapour. A summary of the diagnostics evaluated in this chapter is provided in **Figure 12.2**, which directs the reader towards the appropriate chapter section for further information. Below, we briefly summarize the key findings from this chapter and recommendations for both use of and improvements to reanalysis ozone and water vapour fields.

Key findings of Chapter 4:

- The treatment of ozone and water vapour varies substantially among reanalyses, both in terms of their representation of these species and assimilated observations.
- The latest generation of reanalyses all assimilate satellite total column ozone observations, with some including vertically-resolved measurements.
- Currently none of the reanalyses directly assimilate WV observations in the stratosphere, although they do assimilate temperature and tropospheric humidity observations that can impact their stratospheric water vapour concentrations.
- Comparisons against assimilated observations of total column ozone (TCO) show that reanalyses generally reproduce TCO well in sunlight regions, within ~10 DU (~3 %).
- The lack of TCO observations in polar night, and lack of representation of heterogeneous chemistry in most reanalyses, lead to relatively larger errors in representing TCO in the Antarctic ozone hole.
- From the middle to upper stratosphere, climatological reanalysis ozone profiles are within ±20 % of observations.
- Biases are generally larger (~50 %) for both water vapour and ozone in the upper troposphere and lower stratosphere.
- Significant discontinuities exist in reanalysis water vapour and ozone fields due to transitions in the observing system.

Recommendations from Chapter 4:

- Users should generally use caution when using reanalysis ozone fields for scientific studies and should check that their results are not reanalysis-dependent.
- Reanalysis stratospheric water vapour fields should generally not be used for scientific data analysis (except perhaps for ERA5). Any examination of these fields must account for their inherent limitations and uncertainties.
- In order to improve reanalysis ozone fields, reanalysis centres should work towards improved chemical parameterisations of ozone as well as assimilation of vertically-resolved ozone measurements (*e.g.*, from limb sounders) and measurements in polar night (*e.g.*, from IR nadir sounders).
- In order to improve reanalysis water vapour fields, future efforts should include the collection and assimilation of observational data with sensitivity to stratospheric water vapour, the reduction of reanalysis temperature biases in the TTL, and improvements in the representation of other processes that affect the stratospheric entry mixing ratio.

12.1.3 Chapter 5: Brewer–Dobson Circulation

This chapter presented both a direct comparison of Brewer-Dobson Circulation (BDC)-related dynamical diagnostics from the reanalysis datasets and transport tracer simulations using reanalysis products to drive different offline chemistry-transport models (CTMs). The direct dynamical diagnostics support intercomparison among the reanalyses, whereas the CTM simulations allow comparison against observation-based mean age-of-air (AoA) and stratospheric water vapour distributions, time series, and trends. A summary assessment of representation of the BDC in major reanalyses is provided in **Figure 12.3**, which directs the reader to the appropriate section of the chapter for further information. In the following, we briefly summarize our key findings and recommendations.

Key findings from dynamical diagnostics in Chapter 5:

- The BDC is generally much more consistent and weaker in more recent products compared to their older versions, although there are still significant differences in basic climatological diagnostics for some fields (*e.g.*, shallow branch wave driving, tropical upwelling structure and seasonality, upwelling strength below 70 hPa).

- Dynamical diagnostics show spurious fluctuations in CFSR; this product should thus not be used for long-term trend or interannual variability analyses.

- Estimates of long-term trends (for 1979 - 2016) in tropical upwelling are inconsistent: MERRA-2 and JRA-55 show positive trends, ERA-Interim shows a negative trend, and ERA5 shows no trend.

- Interannual variability and long-term trends in poleward mass transport through the turnaround latitudes ("tropical outwelling") are inconsistent; this suggests that the shallow branch of the BDC is not well constrained, even in the most recent products.

- Latitudinally and vertically resolved trends in residual circulation transit times (RCTTs) show some coherent signatures of a strengthening of the BDC (decreasing RCTTs, especially for the shallow branch), although the afore-mentioned inconsistencies across products also manifest in this diagnostic (especially for the deep branch).

Key findings from transport tracer simulations in Chapter 5:

- Simulations based on more recent reanalyses produce mean AoA in much better agreement with observations than those based on the previous generation of reanalyses (*e.g.*, ERA-Interim *vs.* ERA-40), indicating that reanalysis representations of the BDC have improved. However, significant discrepancies still remain in AoA and tracer distributions among reanalyses, with the spread of AoA obtained using different reanalyses as large as that obtained by using different CCMs.

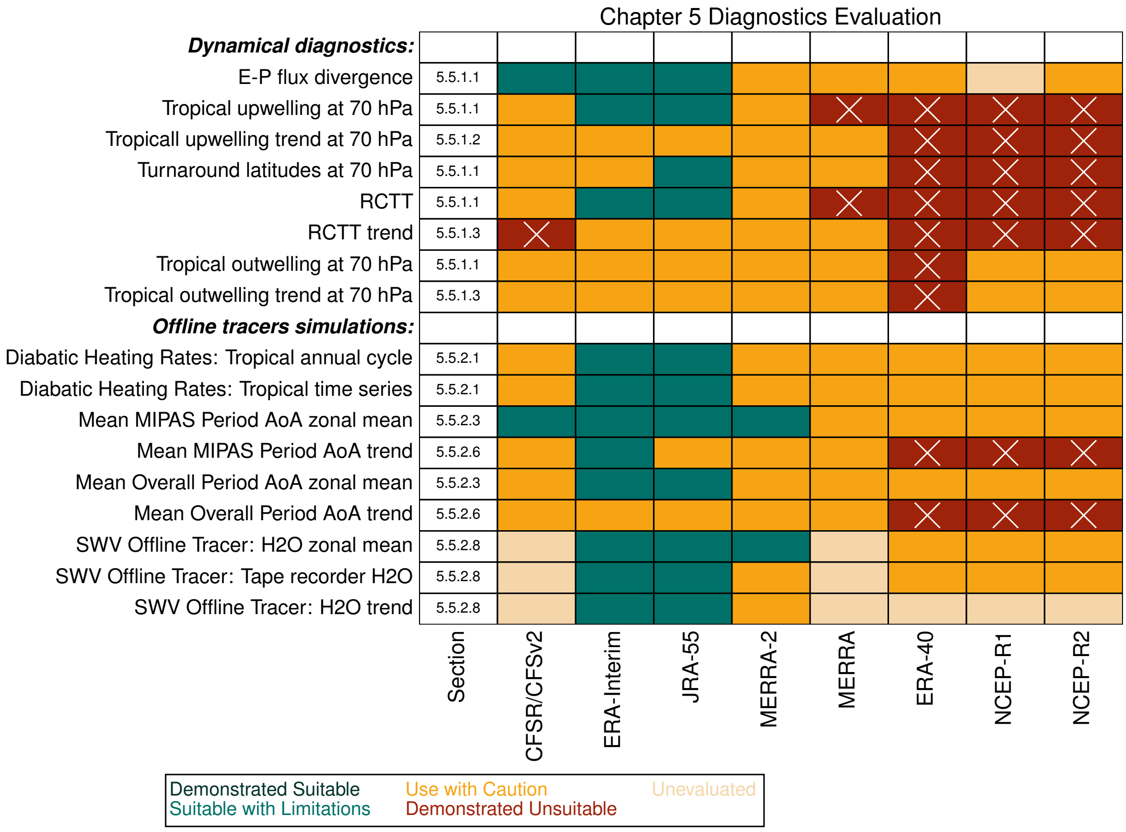
- Differences among reanalysis diabatic heating rates[2] are evident and are a major factor affecting offline simulations of stratospheric tracers using diabatic models. Vertical transport within the tropics is too slow in MERRA and MERRA-2, in agreement with smaller diabatic heating rates compared to the other reanalyses. However, this slower tropical transport is evident in both diabatic and kinematic simulations, indicating that the slower BDC in the GEOS-5 system is not solely attributable to the radiation budget. The RCTT diagnostic also shows longer residence times for MERRA and MERRA-2.

- Our offline simulation results show large spread in the values and signs of AoA trends over 1989 - 2010, depending on the reanalysis and on the region of the stratosphere. For the MIPAS period (2002 - 2012) only ERA-Interim is in good agreement with the observed trends, regardless of the offline model used. A positive trend in the mean AoA in the NH is a robust feature in our studies and is in agreement with other observed phenomena. We emphasize that much investigation is still needed on BDC trends and that these trends should be interpreted with caution regardless of source, as natural variability and changes in the observation system make them highly sensitive to the choice of analysis period.

- Large spread in AoA among reanalyses emerges from two main sources: i) differences among the underlying models used to produce the reanalyses, and ii) the relatively weak constraints on stratospheric transport provided by assimilated observations in reanalyses. AoA diagnostics are affected by many other Earth system phenomena, including the stratospheric QBO signal, ENSO variability, and volcanic eruptions, indicating that improvements in the models and the data assimilation systems can both aid in achieving more accurate BDC representations in future reanalyses.

---

[2] Please note that the diabatic heat budget is not closed in reanalyses. This lack of closure occurs because the data assimilation step can cause changes in temperature that add or remove heat from the system. This analysis increment can be considered as a separate 'diabatic' term in the thermodynamic energy equation, but its application differs amongst reanalyses. Notably, the inclusion of the analysis increment as an additional tendency term in MERRA and MERRA-2 may in turn affect other physical tendency terms produced by the atmospheric model, as the latter are archived during the IAU corrector step rather than the predictor step (see *Chapter 2*, *Section 2.3*). By contrast, tendencies produced by other reanalyses are archived prior to the analysis during the initial forecast/predictor step. The analysis tendency is required to close the budget in either case, but these distinctions should be taken into account when evaluating or interpreting reanalysis diabatic heating products.



	Section	CFSR/CFSv2	ERA-Interim	JRA-55	MERRA-2	MERRA	ERA-40	NCEP-R1	NCEP-R2
Dynamical diagnostics:									
E-P flux divergence	5.5.1.1	🟩	🟩	🟩	🟧	🟧	🟧	⬜	🟧
Tropical upwelling at 70 hPa	5.5.1.1	🟧	🟩	🟩	🟧	☒	☒	☒	☒
Tropicall upwelling trend at 70 hPa	5.5.1.2	🟧	🟧	🟧	🟧	🟧	☒	☒	☒
Turnaround latitudes at 70 hPa	5.5.1.1	🟧	🟧	🟩	🟧	🟧	☒	☒	☒
RCTT	5.5.1.1	🟧	🟩	🟩	🟧	☒	☒	☒	☒
RCTT trend	5.5.1.3	☒	🟧	🟧	🟧	🟧	☒	☒	☒
Tropical outwelling at 70 hPa	5.5.1.1	🟧	🟧	🟧	🟧	🟧	☒	🟧	🟧
Tropical outwelling trend at 70 hPa	5.5.1.3	🟧	🟧	🟧	🟧	🟧	☒	🟧	🟧
Offline tracers simulations:									
Diabatic Heating Rates: Tropical annual cycle	5.5.2.1	🟧	🟩	🟩	🟧	🟧	🟧	🟧	🟧
Diabatic Heating Rates: Tropical time series	5.5.2.1	🟧	🟩	🟩	🟧	🟧	🟧	🟧	🟧
Mean MIPAS Period AoA zonal mean	5.5.2.3	🟩	🟩	🟩	🟩	🟧	🟧	🟧	🟧
Mean MIPAS Period AoA trend	5.5.2.6	🟧	🟩	🟧	🟧	🟧	☒	☒	☒
Mean Overall Period AoA zonal mean	5.5.2.3	🟧	🟩	🟩	🟧	🟧	🟧	🟧	🟧
Mean Overall Period AoA trend	5.5.2.6	🟧	🟧	🟧	🟧	🟧	☒	☒	☒
SWV Offline Tracer: H2O zonal mean	5.5.2.8	⬜	🟩	🟩	🟩	⬜	🟧	🟧	🟧
SWV Offline Tracer: Tape recorder H2O	5.5.2.8	⬜	🟩	🟩	🟧	⬜	🟧	🟧	🟧
SWV Offline Tracer: H2O trend	5.5.2.8	⬜	🟩	🟩	🟧	⬜	⬜	⬜	⬜

Figure 12.3: *(Same as **Figure 5.50**.) A summary of the diagnostics evaluated in Chapter 5: Brewer-Dobson Circulation. The "Section" column at left indicates where in the chapter each diagnostic is described. See the beginning of this chapter for the meaning of the evaluation terms. Note that the score corresponding to "demonstrated suitable" was not assigned to any of the diagnostics listed here, so the darkest green colour does not appear in this table. "E-P flux" = Eliassen-Palm flux; "RCTT" = Residual Circulation Transit Time; "MIPAS" = Michelson Interferometer for Passive Atmospheric Sounding (a satellite instrument); "AoA" = Age of Air; "SWV" = Stratospheric Water Vapour.*

- MERRA-2 shows difficulties in reproducing QBO-related BDC variability before 1995 relative to ERA-Interim and JRA-55. Another feature that is present in MERRA-2 but not in these other two reanalyses is the assimilation of Aura MLS temperatures from 2004 onwards at altitudes above 5hPa. The additional constraints provided by these data can affect stratospheric dynamics, and therefore BDC diagnostics.

Recommendations from Chapter 5:

- MERRA-2 may not be a good option for years before 1995, as it has difficulty reproducing observed QBO variability in stratospheric transport, which also affects its ability to reproduce QBO-related BDC variability.

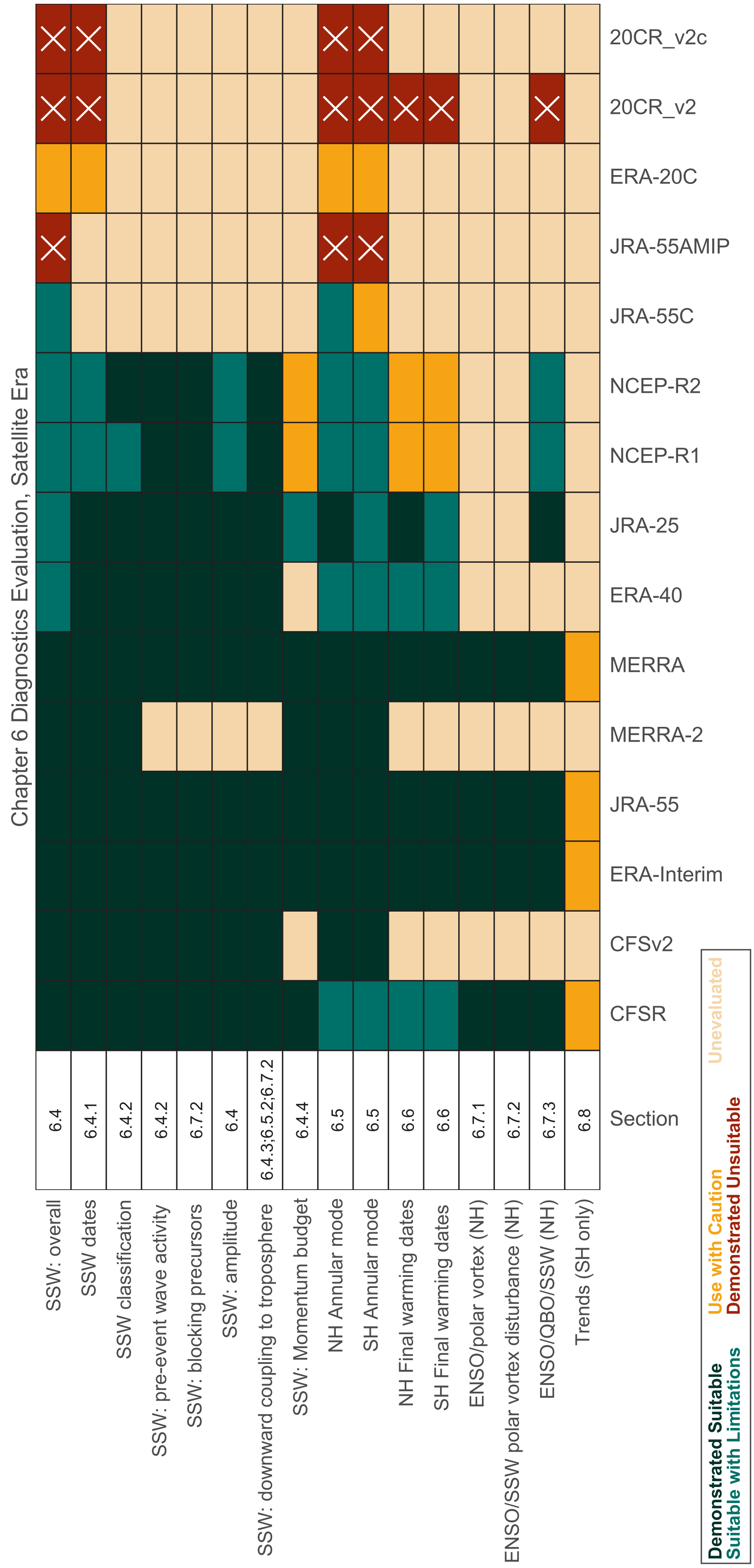
- Among the more recent reanalyses, CFSR has been found to be problematic for BDC studies, especially with respect to interannual variability and long-term trends. Numerous published studies have also shown that older reanalyses like ERA-40, NCEP-NCAR R1, and NCEP-DOE R2 provide unrealistic representations of the BDC and other stratospheric processes. We therefore discourage the use of these older reanalyses for studies of the stratospheric circulation and associated tracer transport.

- Whenever possible we recommend that users not restrict themselves to only one product when conducting studies of the BDC or related transport. In particular, for the period after 2000, comparisons among MERRA-2, JRA-55, and ERA-Interim can help to distinguish robust from non-robust diagnostics.

- We recommend that users work with reanalysis data on model levels for offline simulations and diagnostics related to the shallow branch of the BDC.

- For future reanalyses, we recommend that reanalysis producers: i) provide variable uncertainty information; ii) provide variables at higher vertical resolution, especially within the UTLS region; iii) provide pressure level data at pressures less than 1 hPa (important for RCTT calculations); iv) archive data at higher frequencies; v) archive additional relevant variables (*e.g.*, heating rates) by default.

- The recently released ERA5 includes most of these features, although the resolution around the UTLS is still coarser than desired.



Chapter 6 Diagnostics Evaluation, Satellite Era

	SSW: overall	SSW dates	SSW classification	SSW: pre-event wave activity	SSW: blocking precursors	SSW: amplitude	SSW: downward coupling to troposphere	SSW: Momentum budget	NH Annular mode	SH Annular mode	NH Final warming dates	SH Final warming dates	ENSO/polar vortex (NH)	ENSO/SSW polar vortex disturbance (NH)	ENSO/QBO/SSW (NH)	Trends (SH only)
20CR_v2c	✕	✕	⬜	⬜	⬜	⬜	⬜	⬜	✕	✕	⬜	⬜	⬜	⬜	⬜	⬜
20CR_v2	✕	✕	⬜	⬜	⬜	⬜	⬜	⬜	✕	✕	✕	✕	⬜	⬜	✕	⬜
ERA-20C	🟧	🟧	⬜	⬜	⬜	⬜	⬜	⬜	🟧	🟧	⬜	⬜	⬜	⬜	⬜	⬜
JRA-55AMIP	✕	⬜	⬜	⬜	⬜	⬜	⬜	⬜	✕	✕	⬜	⬜	⬜	⬜	⬜	⬜
JRA-55C	🟦	⬜	⬜	⬜	⬜	⬜	⬜	⬜	🟦	🟧	⬜	⬜	⬜	⬜	⬜	⬜
NCEP-R2	🟦	🟦	🟩	🟩	🟩	🟦	🟩	🟧	🟦	🟦	🟧	🟧	⬜	⬜	🟦	⬜
NCEP-R1	🟦	🟦	🟦	🟩	🟩	🟦	🟩	🟧	🟦	🟦	🟧	🟧	⬜	⬜	🟦	⬜
JRA-25	🟦	🟩	🟩	🟩	🟩	🟩	🟩	🟦	🟩	🟦	🟩	🟦	⬜	⬜	🟩	⬜
ERA-40	🟦	🟩	🟩	🟩	🟩	🟩	🟩	⬜	🟦	🟦	🟦	🟦	⬜	⬜	⬜	⬜
MERRA	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟧
MERRA-2	🟩	🟩	🟩	⬜	⬜	⬜	⬜	🟩	🟩	🟩	⬜	⬜	⬜	⬜	⬜	⬜
JRA-55	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟧
ERA-Interim	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟧
CFSv2	🟩	🟩	🟩	🟩	🟩	🟩	🟩	⬜	🟩	🟩	⬜	⬜	⬜	⬜	⬜	⬜
CFSR	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟦	🟦	🟦	🟦	🟩	🟩	🟩	🟧
Section	6.4	6.4.1	6.4.2	6.4.2	6.7.2	6.4	6.4.3;6.5.2;6.7.2	6.4.4	6.5	6.5	6.6	6.6	6.7.1	6.7.2	6.7.3	6.8

🟩 **Demonstrated Suitable**
🟦 **Suitable with Limitations**
🟧 **Use with Caution**
✕ **Demonstrated Unsuitable**
⬜ **Unevaluated**

Figure 12.4: *(Same as* ***Figure 6.23****.) Evaluation of reanalyses during the satellite era (1979 onward) based on diagnostics computed for Chapter 6: Extratropical Stratosphere-Troposphere Coupling. The "Section" column at left indicates where in the chapter each diagnostic is described. See the beginning of this chapter for the meaning of the evaluation terms. "SSW" = Sudden Stratospheric Warmings; "NH" = Northern Hemisphere; "SH" = Southern Hemisphere; "ENSO" = El Niño–Southern Oscillation; "QBO" = Quasi-Biennial Oscillation.*

- Further model development will be required to improve the representation of the BDC in future reanalyses. Aspects that require particular attention are: i) gravity wave drag parameterisations; ii) representations of radiative gases and aerosols in the stratosphere; iii) cloud and convection parameterisations, especially in tropical latitudes; iv) assimilation of stratospheric winds; v) model vertical resolution in the UTLS; and vi) extension of the vertical range to incorporate mesospheric processes.

- Sustained long-term observation platforms are required to monitor changes in the strength and structure of the BDC, to keep evaluating how well current and future reanalyses represent major stratospheric circulation patterns. Therefore, we strongly recommend the creation and sustained support of such observation platforms, and that they operate long enough to cover time scales relevant to the evolution and trends of the BDC.

12.1.4 Chapter 6: Extratropical Stratosphere–Troposphere Coupling

Atmospheric reanalyses are vital for evaluating stratosphere-troposphere coupling due to the lack of direct observations of the large-scale atmospheric circulation. In this chapter, we examined the representation of coupling between the troposphere and stratospheric polar vortices across the reanalyses. We assessed the reanalyses in terms of their internal consistency and in terms of their consistency with one another. Summary assessments of key stratosphere-troposphere coupling diagnostics are provided in **Figure 12.4** for the satellite era (1979 and later) and **Figure 12.5** for the pre-satellite era (1958 - 1978). Both figures direct the reader to the appropriate section of the chapter for further information. In the following, we briefly summarize key findings and recommendations based on our evaluation.

Key findings of Chapter 6:

- In the satellite era (1979 - onward), the representation of large scale stratosphere-troposphere circulation is very consistent across all full-input reanalyses. On synoptic scales, the more recent reanalyses (ERA-Interim, JRA-55, MERRA, and MERRA-2, and to a slightly lesser extent, CFSR/CFSv2) become more clearly superior.

- Our ability to assess and understand stratosphere-troposphere coupling is primarily limited by sampling uncertainty, that is, by the comparatively large natural variability of the circulation relative to the length of the satellite record. As an example, various efforts have sought to characterize the break-down of the polar vortex during a Sudden Stratospheric Warmings (SSW) as a split or displacement event. Methodological differences among the classifications proposed in the literature, however, result in a partial agreement (for two-thirds of SSW events). In contrast, applying the same definition to different reanalyses yields nearly identical results.

- Although measures of stratosphere-troposphere coupling determined from earlier reanalyses are generally not statistically distinct from results obtained with a more recent reanalysis, the more recent products show demonstrable improvement, particularly with respect to internal consistency (*e.g.*, the momentum budget) and at higher levels (10 hPa and above).

- Reanalysis datasets broadly agree on trends in the austral polar vortex related to ozone depletion since 1979. In contrast, there are no discernible trends in Northern Hemisphere polar vortex variability.

- Pre-satellite era reanalyses (1958 - 1978) appear to be of good quality in the Northern Hemisphere, and therefore can be used to reduce sampling uncertainty in measures of stratosphere-troposphere coupling by approximately 20 %. We emphasize that this represents a more significant reduction in uncertainty than achieved by shifting from an earlier generation reanalysis to a more recent reanalysis.

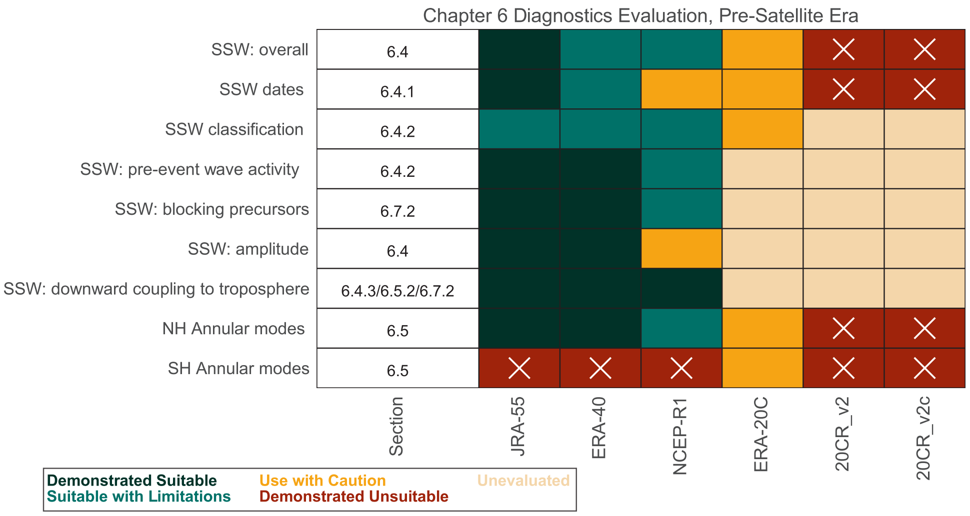
- Pre-satellite era reanalyses of the Southern Hemisphere are generally of poor quality, and can only be used to reduce sampling uncertainty with great caution.

- A conventional-input reanalysis of the Northern Hemisphere (JRA-55C) matches full-input reanalyses well up to 10 hPa, supporting the validity of pre-satellite reanalysis products in this hemisphere. JRA-55C's representation of the Southern Hemisphere is not as accurate, suggesting that satellite measurements are more critical in this hemisphere due to the reduced density of conventional observations.

- Surface-input reanalyses have also been evaluated. ERA-20C captures not only the correct statistical climatology of the Northern Hemisphere stratospheric polar vortex, but also much of its actual variability (correctly representing the timing of about half of observed SSWs). This suggests it may be suitable for exploring low-frequency variability of the stratosphere-troposphere coupled system. The representation of the stratospheric vortex in NOAA 20CR v2/v2c, however, is demonstrably poor.

Recommendations from Chapter 6:

- We recommend the use of more recent reanalysis products. As a matter of best practice, we urge all users to avoid the use of earlier reanalyses unless the project requires the use of an older product, and special care is taken to justify that the older product is otherwise consistent with more recent reanalyses. In particular, we note for users that modern reanalyses can be obtained, in addition to their native high-resolution grids, at a coarser resolution that is comparable to that of earlier reanalyses and thus more manageable in size, but which still captures the best representation of the large-scale circulation.



Chapter 6 Diagnostics Evaluation, Pre-Satellite Era

	Section	JRA-55	ERA-40	NCEP-R1	ERA-20C	20CR_v2	20CR_v2c
SSW: overall	6.4	■	▨	▨	🟧	☒	☒
SSW dates	6.4.1	■	▨	🟧	🟧	☒	☒
SSW classification	6.4.2	▨	▨	▨	🟧	□	□
SSW: pre-event wave activity	6.4.2	■	■	▨	□	□	□
SSW: blocking precursors	6.7.2	■	■	▨	□	□	□
SSW: amplitude	6.4	■	■	🟧	□	□	□
SSW: downward coupling to troposphere	6.4.3/6.5.2/6.7.2	■	■	■	□	□	□
NH Annular modes	6.5	■	■	▨	🟧	☒	☒
SH Annular modes	6.5	☒	☒	☒	🟧	☒	☒

■ Demonstrated Suitable
▨ Suitable with Limitations
🟧 Use with Caution
☒ Demonstrated Unsuitable
□ Unevaluated

Figure 12.5: *(Same as **Figure 6.24**.) Evaluation of reanalyses during the pre-satellite era (1958 - 1978) based on diagnostics computed for Chapter 6: Extratropical Stratosphere-Troposphere Coupling. The "Section" column at left indicates where in the chapter each diagnostic is described. See the beginning of this chapter for the meaning of the evaluation terms. "SSW" = Sudden Stratospheric Warmings; "NH" = Northern Hemisphere; "SH" = Southern Hemisphere.*

- The consistency of trends associated with the Antarctic ozone hole (for the period 1979 forward) suggest that reanalyses may be reliably capturing the influence of stratospheric ozone loss. One must exercise great caution in the interpretation of trends in the reanalyses, however, as they can be spuriously caused by changes in the observations assimilated over time, an issue that could systematically affect all products. Additional support from direct observations and/or understanding of the mechanism(s) help build confidence in trends found in the reanalyses.

- When an extended record is needed to reduce sampling uncertainty, we recommend the use of pre-satellite era reanalyses (1958 - 1978) in the Northern Hemisphere, but caution against their use in the Southern Hemisphere.

- Due to significant biases in the mean state and variability of the polar vortex in the NOAA 20CR surface-input reanalysis, we do not recommend it for the purpose of investigating stratosphere-troposphere coupling.

- ERA-20C may be suitable, with caution, for exploring the low-frequency variability of the stratosphere-troposphere coupled system.

- As our ability to quantify the large scale coupling between the stratosphere and troposphere is primarily limited by sampling uncertainty, we recommend that future reanalysis products extend their analysis prior to the satellite era.

12.1.5 Chapter 7: Extratropical Upper Troposphere and Lower Stratosphere (ExUTLS)

In this chapter, we have evaluated diagnostics that are critical to understanding ExUTLS dynamical and transport processes, including the extratropical tropopause; upper tropospheric (UT) jet streams; mixing and transport diagnostics; and ozone distributions and evolution. Because representing these processes requires high resolution, we focus on recent full-input reanalyses, including MERRA, MERRA-2, ERA-Interim, JRA-55, and CFSR/CFSv2, with the conventional input JRA-55C also included for a few diagnostics. **Figure 12.6** summarizes the results for the main diagnostics evaluated in this chapter, and directs the reader to the appropriate section of the chapter for further information. Because most of the diagnostics evaluated in *Chapter 7* cannot be verified using direct observations, there are very few cases where we can rate the reanalyses as "demonstrated suitable".

We summarize our key findings and recommendations below.

Key Findings of Chapter 7:

- The reanalyses evaluated here agree well on the placement of the lapse-rate tropopause, both with each other and with data from high-resolution radiosonde observations. CFSR/CFSv2 shows the smallest errors with respect to radiosonde-based lapse-rate tropopause data.

- Long-term trends in tropopause characteristics are in broad agreement both among the reanalyses and with observations, except for CFSR/CFSv2.

- The representation of multiple lapse-rate tropopause altitudes, which indicate lateral stratosphere-troposphere exchange (STE) events between the tropical UT and extratropical LS, is highly dependent on the vertical grid resolution of the reanalysis. CFSR/CFSv2 has the highest frequency of multiple tropopauses, as well as the highest ExUTLS resolution among the reanalyses evaluated here.

- Using pressure and model-level versions of CFSR/CFSv2, we show that the coarser vertical resolution of the pressure-level fields makes them unsuitable for identifying tropopause locations, especially in multiple-tropopause situations.

- JRA-55C is unsuitable for identifying multiple tropopauses because of its inability to qualitatively reproduce the distributions in SH high latitudes.

- Despite a general under-representation of multiple tropopause frequency compared to observations, most modern reanalyses reproduce the pattern and sign of observed long-term trends.

- The reanalyses show good overall agreement in representing the climatologies of UT jets and the sub-vortex jet in the lowermost stratosphere.

- Robust trends in UT jets (latitude, altitude, and windspeed) are limited to particular longitude regions and seasons. Disagreement among the reanalyses is most common for the SH jets; in particular, MERRA-2 and/or CFSR/CFSv2 sometimes differ from the other reanalyses even in the sign of the SH jet latitude trend.

- Kinematic STE is in broad agreement among the reanalyses, with some important differences in the magnitudes and long-term changes of troposphere-to-stratosphere transport and stratosphere-to-troposphere transport. Transport estimates are sensitive to the choice of vertical coordinate (*i.e.*, diabatic vs. kinematic) and the period analyzed.

- Mixing diagnostics including effective diffusivity and PV gradients as a function of equivalent latitude (EqL) show generally good agreement in both climatological seasonal cycles and interannual variability.

- Mass flux across the 380 K isentropic surface agrees well among MERRA-2, ERA-Interim, and JRA-55, but CFSR/CFSv2 shows inconsistencies in the seasonal cycle.

- Climatological ozone distributions and seasonal cycles show good qualitative agreement. Given large differences in the ozone products assimilated and the methods of assimilating them, this points to good representations of the dynamics in the UTLS, where ozone changes are primarily driven by dynamical and transport processes.

- Reanalysis ozone fields mapped in EqL generally reproduce at least qualitatively the interannual variability in MLS-observed ozone, but ERA-Interim shows several step function changes that are related to changes in the versions of MLS ozone assimilated. For example, large biases in ERA-Interim UTLS ozone arise in mid-2009 through 2012 owing to the use of an early version of MLS near real time data.

Recommendations from Chapter 7:

- Only the recent high-resolution reanalyses (MERRA-2, ERA-Interim, JRA-55, and CFSR/CFSv2 are such reanalyses evaluated herein) are suitable for ExUTLS dynamical and transport studies. Dynamical diagnostics derived from these reanalyses indicate that they are all suitable for use in such studies with some limitations. Earlier reanalyses (*e.g.*, ERA-40, NCEP-NCAR R1, and NCEP-DOE R2) are not suitable for detailed UTLS studies and are not evaluated here.

- A few diagnostics (*e.g.*, effective diffusivity in CFSR/CFSv2; ozone in ERA-Interim) show substantial discontinuities when assessed over many years, and thus should be used with greatest caution and awareness.

- Because many diagnostics in this chapter cannot be directly compared with observational data, it is important that ExUTLS studies use multiple reanalyses and assess agreement among them whenever possible.

- For diagnostics that cannot be directly compared with data, and in light of similar changes in input data, agreement among the reanalyses should be regarded as a necessary but by no means sufficient condition for robustness of trends.

- As is the case for diagnostics described in other chapters (*e.g.*, *Chapter 10*), differences between the PV fields arising from differing products provided by the reanalysis centres add to uncertainties in the evaluations. It would be helpful in the future for all reanalysis centres to provide PV on the model grids.

- The results from reanalyses assimilating MLS ozone (which has relatively high vertical resolution compared to other ozone profilers currently used) show promise for future improvements. More attention to consistently assimilating high-resolution ozone observations in future reanalyses would be extremely beneficial to understanding the processes controlling ozone in this region, where it is of great importance to the radiative balance.

- Future work is needed to better elucidate the role of various elements of model design in producing observed differences in tropopause location and characteristics (*e.g.*, through idealized simulations with the core models of each reanalysis).

- In the future, the accuracy of tropopause identifications in reanalyses should improve as the vertical grid spacing decreases. These diagnostics should be evaluated in forthcoming reanalyses (most immediately in ERA5) and the impacts of these improvements on estimates of STE and their long-term changes should be explored.

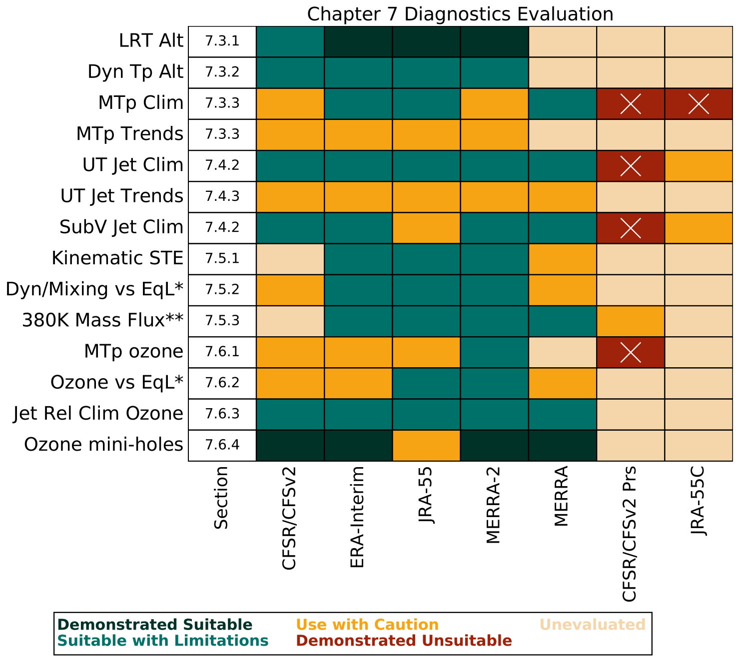


Figure 12.6: *(Same as* ***Figure 7.36****.) A summary of the diagnostics evaluated in Chapter 7: Extratropical Upper Troposphere and Lower Stratosphere (ExUTLS). The "Section" column at left indicates where in the chapter each diagnostic is described. See the beginning of this chapter for the meaning of the evaluation terms. Because analyses as a function of equivalent latitude (EqL; marked by *) depend critically on PV (which is used to compute EqL), reanalyses where we have concerns about the PV fields are rated "use with caution" even in the absence of obvious "red flags". "CFSR/CFSv2 Prs" indicates CFSR/CFSv2 was used as interpolated to standard pressure levels, and the 380 K mass flux analysis (marked by **) was done using pressure level data for all reanalyses. All other diagnostics were calculated using model level data for all reanalyses. "LRT" = lapse-rate tropopause; "Alt" = altitude; "Dyn" = dynamical; "Tp" = tropopause; "MTp" = multiple tropopause; "UT" = upper troposphere/tropospheric; "Clim" = climatology; "SubV" = subvortex; "STE" = stratosphere-troposphere exchange; "Jet Rel" = in coordinates relative to the subtropical jet core location.*

- The accuracy of transport estimates from reanalyses is largely unknown, since global estimates of transport from observing systems are not available and the outcomes are sensitive to the input fields and methods used. Comparison of transport calculations using reanalysis wind fields and trace gas observations is one path to examine the accuracy of transport in reanalyses.

- If possible, errors in transport calculations should be increasingly gleaned from comparison of trajectory calculations driven by the reanalysis winds to long-duration balloon observations. However, such observations are infrequent and sometimes assimilated into the reanalysis, which limits their utility for validation studies.

- Given known errors in trajectory and other transport calculations that arise from coarse temporal resolution of input wind fields, more frequent 3D wind field outputs are desired from future reanalyses. Such wind fields, which are already available for ERA5, will allow for improved understanding of transport and STE.

- Increased horizontal and vertical grid resolution will also be beneficial for reducing errors in transport calculations and enabling analysis of processes at smaller scales.

12.1.6 Chapter 8: Tropical Tropopause Layer (TTL)

In this chapter, we have investigated the extent to which reanalysis data sets reproduce key characteristics of the TTL, including the cold point and lapse rate tropopause, the vertical structure and distribution of clouds within the TTL, basic dynamical processes and circulation patterns, transport statistics and residence times derived from trajectory simulations, equatorial wave activity, and long-term changes in the width of the tropical belt. We have also evaluated how key differences in reanalysis performance within the TTL impact upon regional and seasonal aspects of the South Asian Summer Monsoon (SASM) anticyclone. Summary assessments of reanalysis products in the TTL are provided in **Figure 12.7** for the global tropics and in **Figure 12.8** for the SASM. Key findings and recommendations from this chapter are outlined below.

Key findings of Chapter 8:

- Advances in reanalysis and observational systems over recent years have led to a clear improvement in TTL reanalysis products over time. In particular, the reanalyses ERA-Interim, ERA5, MERRA-2, CFSR, and JRA-55 show very good agreement after 2002 in terms of the vertical TTL temperature profile, meridional tropopause structure, and interannual variability.

Long-term temperature trends from reanalyses and adjusted radiosonde data indicate significant cooling in the upper TTL (above the cold point).

- While climatological TTL temperatures from reanalyses agree very well with observations with relatively small low biases, the cold point and lapse rate tropopause show warm biases, most likely related to the fact that the discrete values corresponding to reanalysis model levels are unable to reproduce the observed minimum temperature as recorded in a near-continuous profile.

- Cloud fields in the tropical UTLS vary greatly in both magnitude and vertical distribution across reanalyses. Differences in cloud fraction and cloud water content impact the radiation budget both at the top-of-atmosphere and within the UTLS, and the effects of differences in cloud and convection parameterizations can be identified in vertical profiles of temperature and humidity in the tropical troposphere.

- There are large differences among reanalysis diabatic heating rate products[3] within the TTL, which are known to influence transport statistics and rates of ascent in trajectory simulations of cross-tropopause transport in this region. Differences among reanalysis diabatic heating rates in the tropical UTLS are not limited to any one component: longwave, shortwave, and non-radiative components all show substantial discrepancies.

- Lagrangian transport studies demonstrate large differences in reanalysis temperatures at the dehydration point and in TTL residence times. However, the data sets agree on the spatial distribution of dehydration locations and produce roughly similar distributions, seasonal cycles, and interannual variations of TTL residence time.

- Equatorial wave activity and corresponding temperature anomaly patterns at 100 hPa are similar among the reanalyses, including the characteristic horseshoe-shaped structures that resemble the stationary wave response to tropical heating. However, the strength of the wave activities, their spectral magnitudes, and the intensity of temperature response differ among the reanalyses, with the latter differences depending on the aspects of the dynamical model and/or assimilation system.

- Metrics of the width of the TTL based on the zonally-resolved subtropical jet and tropopause break show robust changes in only a few regions and seasons and poor agreement of the resulting zonal-mean annual-mean values. The diagnostics based on the zonal-mean subtropical jet and tropopause break, on the other hand, suggest stronger trends in the width of the TTL than their zonally-resolved counterparts. Overall, the two subtropical jet diagnostics are more consistent than the two tropopause break diagnostics, possibly related to smoother variations in the zonal wind field relative to the tropopause break.

- Modern reanalyses agree well regarding the climatological position and evolution of area extent and moments of the SASM anticyclone, although there are notable differences in the distribution of SASM anticyclone centre locations. All of the reanalyses indicate slightly higher CPT temperatures and lower CPT heights in the SASM anticyclone compared to GNSS-RO satellite observations.

- Distributions of ozone volume mixing ratios within the SASM anticyclone are qualitatively consistent among reanalyses and broadly consistent with observations. However, none of the evaluated reanalyses are able to reproduce the low ozone mixing ratios within the SASM anticyclone.

- Cloud properties, convection, radiative heating, and omega fields for the SASM UTLS differ significantly among reanalyses on a regional scale as these properties are only weakly constrained by assimilated observations. These differences impact derived transport processes in the UTLS, and residence times based on diabatic Lagrangian transport calculations reveal large differences.

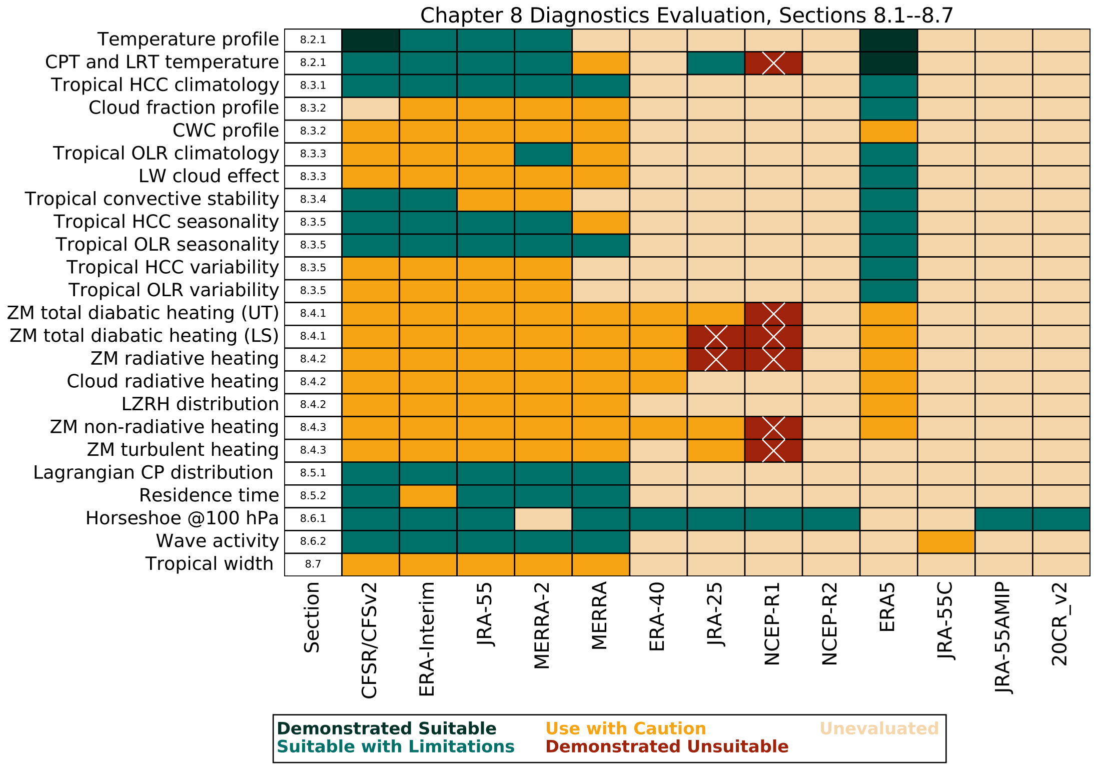
Recommendations from Chapter 8:

- In the TTL, temperature on native model levels should be used rather than the standard pressure-surface data sets. Various diagnostics such as the cold point and lapse rate tropopause and the analysis of equatorial waves are demonstrably improved when model-level data are used. For a more realistic representation of the tropical tropopause levels, data sets that combine low temperature biases with high vertical resolution should be used.

- Long-term drifts in high cloud fraction, OLR, and LWCRE are present in almost all reanalyses, and often disagree in terms of sign, timing, or magnitude. These products should generally not be used for trend or time series analysis without independent verification. Among the reanalyses, ERA5 shows greater stability in time and stronger correlations with observed variability for these cloud and radiation metrics and may therefore offer a more reliable characterization of long-term variations in related metrics relative to earlier reanalyses.

---

[3] See the footnote on diabatic heating rates in reanalyses in *Section 12.1.3.*



Chapter 8 Diagnostics Evaluation, Sections 8.1--8.7

	Section	CFSR/CFSv2	ERA-Interim	JRA-55	MERRA-2	MERRA	ERA-40	JRA-25	NCEP-R1	NCEP-R2	ERA5	JRA-55C	JRA-55AMIP	20CR_v2
Temperature profile	8.2.1	⬛	🟩	🟩	🟩	⬜	⬜	⬜	⬜	⬜	⬛	⬜	⬜	⬜
CPT and LRT temperature	8.2.1	🟩	🟩	🟩	🟩	🟧	⬜	🟩	☒	⬜	⬛	⬜	⬜	⬜
Tropical HCC climatology	8.3.1	🟩	🟩	🟩	🟩	🟩	⬜	⬜	⬜	⬜	🟩	⬜	⬜	⬜
Cloud fraction profile	8.3.2	⬜	🟧	🟧	🟧	🟧	⬜	⬜	⬜	⬜	🟩	⬜	⬜	⬜
CWC profile	8.3.2	🟧	🟧	🟧	🟧	🟧	⬜	⬜	⬜	⬜	🟧	⬜	⬜	⬜
Tropical OLR climatology	8.3.3	🟧	🟧	🟧	🟩	🟧	⬜	⬜	⬜	⬜	🟩	⬜	⬜	⬜
LW cloud effect	8.3.3	🟧	🟧	🟧	🟧	🟧	⬜	⬜	⬜	⬜	🟩	⬜	⬜	⬜
Tropical convective stability	8.3.4	🟩	🟩	🟧	🟧	⬜	⬜	⬜	⬜	⬜	🟩	⬜	⬜	⬜
Tropical HCC seasonality	8.3.5	🟩	🟩	🟩	🟩	🟧	⬜	⬜	⬜	⬜	🟩	⬜	⬜	⬜
Tropical OLR seasonality	8.3.5	🟩	🟩	🟩	🟩	🟩	⬜	⬜	⬜	⬜	🟩	⬜	⬜	⬜
Tropical HCC variability	8.3.5	🟧	🟧	🟧	🟧	⬜	⬜	⬜	⬜	⬜	🟩	⬜	⬜	⬜
Tropical OLR variability	8.3.5	🟧	🟧	🟧	🟧	⬜	⬜	⬜	⬜	⬜	🟩	⬜	⬜	⬜
ZM total diabatic heating (UT)	8.4.1	🟧	🟧	🟧	🟧	🟧	🟧	🟧	☒	⬜	🟧	⬜	⬜	⬜
ZM total diabatic heating (LS)	8.4.1	🟧	🟧	🟧	🟧	🟧	🟧	☒	☒	⬜	🟧	⬜	⬜	⬜
ZM radiative heating	8.4.2	🟧	🟧	🟧	🟧	🟧	🟧	☒	☒	⬜	🟧	⬜	⬜	⬜
Cloud radiative heating	8.4.2	🟧	🟧	🟧	🟧	🟧	🟧	⬜	⬜	⬜	🟧	⬜	⬜	⬜
LZRH distribution	8.4.2	🟧	🟧	🟧	🟧	🟧	⬜	⬜	⬜	⬜	🟧	⬜	⬜	⬜
ZM non-radiative heating	8.4.3	🟧	🟧	🟧	🟧	🟧	🟧	🟧	☒	⬜	🟧	⬜	⬜	⬜
ZM turbulent heating	8.4.3	🟧	🟧	🟧	🟧	🟧	⬜	🟧	☒	⬜	⬜	⬜	⬜	⬜
Lagrangian CP distribution	8.5.1	🟩	🟩	🟩	🟩	🟩	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜
Residence time	8.5.2	🟩	🟧	🟩	🟩	🟩	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜
Horseshoe @100 hPa	8.6.1	🟩	🟩	🟩	⬜	🟩	🟩	🟩	🟩	🟩	⬜	⬜	🟩	🟩
Wave activity	8.6.2	🟩	🟩	🟩	🟩	🟩	⬜	⬜	⬜	⬜	⬜	🟧	⬜	⬜
Tropical width	8.7	🟧	🟧	🟧	🟧	🟧	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜

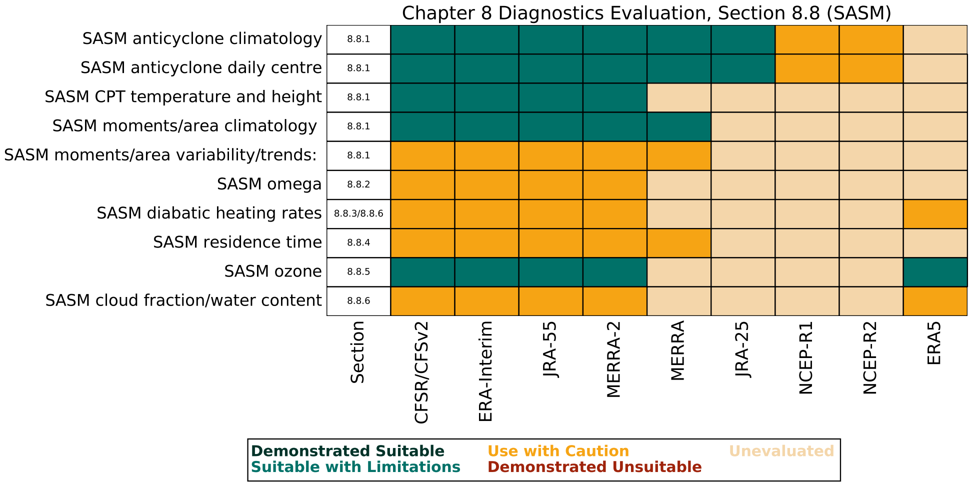
⬛ Demonstrated Suitable
🟩 Suitable with Limitations
🟧 Use with Caution
☒ Demonstrated Unsuitable
⬜ Unevaluated

Figure 12.7: *(Same as **Figure 8.72**, top.) A summary of the diagnostics evaluated in Sections 8.2-8.7 of Chapter 8: Tropical Tropopause Layer (TTL). The "Section" column at left indicates where in the chapter each diagnostic is described. See the beginning of this chapter for the meaning of the evaluation terms. "CPT" = cold point tropopause; "LRT" = lapse rate tropopause; "HCC" = high cloud cover fraction; "CWC" = cloud water content; "OLR" = outgoing longwave radiation; "LW" = longwave; "ZM" = zonal mean; "UT"=Upper Troposphere; "LS"=Lower Stratosphere; "LZRH" = level of zero net radiative heating; "CP" = cold point.*

- Given large differences in reanalysis diabatic heating products and related metrics within the tropical UTLS, researchers using these fields to drive or nudge model simulations of this region should use multiple reanalyses whenever possible.

- When applying metrics of tropical width based on the subtropical jet or tropopause break, it is recommended to use multiple reanalyses and to be aware of the caveat that the zonal-mean diagnostics suggest stronger trends than their zonally-resolved counterparts.

- For analyses involving the SASM anticyclone it is recommended to use more recent reanalyses. In particular, researchers are encouraged to avoid NCEP-NCAR R1 and NCEP-DOE R2 data sets and the geopotential height field of the MERRA-2-ANA pressure-level data when possible.



Chapter 8 Diagnostics Evaluation, Section 8.8 (SASM)

	Section	CFSR/CFSv2	ERA-Interim	JRA-55	MERRA-2	MERRA	JRA-25	NCEP-R1	NCEP-R2	ERA5
SASM anticyclone climatology	8.8.1	🟩	🟩	🟩	🟩	🟩	🟩	🟧	🟧	⬜
SASM anticyclone daily centre	8.8.1	🟩	🟩	🟩	🟩	🟩	🟩	🟧	🟧	⬜
SASM CPT temperature and height	8.8.1	🟩	🟩	🟩	🟩	⬜	⬜	⬜	⬜	⬜
SASM moments/area climatology	8.8.1	🟩	🟩	🟩	🟩	🟩	⬜	⬜	⬜	⬜
SASM moments/area variability/trends:	8.8.1	🟧	🟧	🟧	🟧	🟧	⬜	⬜	⬜	⬜
SASM omega	8.8.2	🟧	🟧	🟧	🟧	⬜	⬜	⬜	⬜	⬜
SASM diabatic heating rates	8.8.3/8.8.6	🟧	🟧	🟧	🟧	⬜	⬜	⬜	⬜	🟧
SASM residence time	8.8.4	🟧	🟧	🟧	🟧	🟧	⬜	⬜	⬜	⬜
SASM ozone	8.8.5	🟩	🟩	🟩	🟩	⬜	⬜	⬜	⬜	🟩
SASM cloud fraction/water content	8.8.6	🟧	🟧	🟧	🟧	⬜	⬜	⬜	⬜	🟧

⬛ Demonstrated Suitable
🟩 Suitable with Limitations
🟧 Use with Caution
☒ Demonstrated Unsuitable
⬜ Unevaluated

Figure 12.8: *(Same as **Figure 8.72**, bottom.) A summary of the diagnostics evaluated in Section 8.8 of Chapter 8: South Asian Summer Monsoon (SASM). The "Section" column at left indicates where in Section 8.8 each diagnostic is described. See the beginning of this chapter for the meaning of the evaluation terms. "CPT" = cold point tropopause.*

- Transport simulations for the SASM domain that use diabatic heating rates to represent vertical motion should use multiple reanalyses if possible and carefully consider the representation of convective sources to the TTL. MERRA-2 diabatic heating rates should only be used at 370 K potential temperature level and above.

- Ozone in the UTLS above the SASM should be carefully validated against observations, and cloud and radiative heating should be used with caution for all reanalyses.

12.1.7 Chapter 9: Quasi-Biennial Oscillation (QBO)

In this chapter, we have investigated the representation of the quasi-biennial oscillation (QBO) and tropical stratospheric variability in atmospheric reanalyses. An assessment of key QBO-related diagnostics is provided in **Figure 12.9**, which includes cross-references for finding further information in the chapter text. Here we provide a concise summary along with recommendations on which reanalyses are appropriate to use for various diagnostics of the QBO and tropical stratospheric variability.

Key findings of Chapter 9:

- Reanalyses broadly agree with FUB winds on the evolution of the zonal-wind QBO apart from the older NCEP reanalyses (NCEP-NCAR R1 and NCEP-DOE R2), although even these adequately reproduce the phase of the QBO. The main error in NCEP-NCAR R1 and NCEP-DOE R2 is that the QBO wind amplitude is substantially underestimated (by a factor of 2 - 4, depending on the altitude considered).

- Inter-reanalysis spread in QBO winds has decreased in recent years, consistent with increasing observations to constrain the reanalyses. However, differences between JRA-55 and JRA-55C show no long-term trend, indicating that the increased satellite data assimilated into JRA-55 over the 1973 - 2012 period does not substantially affect the QBO winds. This suggests that satellite observations are less important than conventional observations for constraining the QBO.

- Most inter-reanalysis spread in QBO winds occurs during QBO phase transitions, especially the QBO-W (westerly) onset which is often delayed by ~ 1 - 2 months compared with FUB winds. These onsets are also delayed when compared with the MERRA-2 reanalysis, which uses a forecast model that spontaneously generates a QBO. Hence, we attribute the delays to lack of sufficiently strong westerly momentum deposition in the tropical stratosphere, that can only be provided by wave drag.

- There is substantial inter-reanalysis spread in strength and spatial structure of zonal winds in the tropical upper troposphere and tropopause region (both zonal-mean and zonally-varying components). This has implications for modelling tropical wave propagation (*i.e.*, how the background winds filter upward propagation of waves that force the QBO and SAO, including parameterized gravity waves). Small changes in wave filtering at lower altitudes can have substantial effects on wave forcing at higher altitudes.

- There is uncertainty in how much zonal asymmetry is present in the QBO, especially at 70 hPa, given that assimilation of winds in the tropics is dominated by the Singapore radiosondes. Inter-reanalysis spread is greatest over the oceans where there is a lack of radiosonde observations. Inter-reanalysis spread has reduced in recent years but spatial patterns remain unchanged, especially at 70 hPa where the flow is less zonally symmetric. QBO-related vertical velocity anomalies have comparable magnitude to the background vertical velocity, though the magnitudes of both vary among the reanalyses.

- Reanalysis QBO temperature anomaly evolutions compare well with sonde and GNSS-RO observations (all reanalyses considered here assimilate radiosondes, and the four recent 'full-input' reanalyses (ERA-Interim, CFSR, JRA-55, MERRA-2) assimilate GNSS-RO data, albeit over slightly different periods). Peak-to-peak QBO zonal-mean temperature variations are ~ 2 K at 70 hPa and ~ 1 K near the tropical tropopause (100 hPa), corresponding to 25 - 30 % and 15 - 20 % the size of the annual cycle, respectively. Zonal asymmetries are also evident, with QBO amplitude in the Indonesian region roughly 30 % larger than the zonal-mean amplitude. Comparison with GNSS-RO, which are spatially homogeneous, suggests this is a real feature rather than an artefact of the strong influence of the Singapore observations. This may have implications for QBO influences on convection and precipitation.

- There is good agreement on the relative contributions of the various tropical waves to forcing the QBO. The greatest inter-reanalysis spread is in the Kelvin wave contribution during the descending QBO-W phase. There is significant natural variability (*i.e.*, from one QBO phase to the next) in the various contributions. The vertical advection term differs widely among reanalyses, including in its sign, consistent with large inter-reanalysis differences in vertical velocity.

- Although assimilation of satellite observations does not have a major impact on the QBO wind evolution (as noted above) it nevertheless has an indirect impact via improved representation of different components of the waves that force the QBO, which may in turn contribute to improvements in details such as the spread in the timing of QBO phase changes referred to above. There is clear evidence that representations of tropical waves changed after introduction of the AMSU satellite observations in ~ 1998. Assuming that the observations are more accurate in the latter period, we recommend that the more recent data be used for studies of wave diagnostics.

- There are clear differences in wave characteristics when derived on model versus pressure surfaces. They are qualitatively similar, but for quantitative results model levels are better. Comparison of wave characteristics with satellite observations (HIRDLS, SABER, COSMIC, and AIRS) shows consistency between the reanalyses and high correlations in the tropical lower stratosphere with all observations except AIRS. Correlations with HIRDLS and SABER are notable because these observations are not assimilated by any of the reanalyses and thus provide independent validation. Reanalysis momentum fluxes in the lower tropical stratosphere correlate well with HIRDLS but less well with SABER.

- There is good inter-reanalysis agreement on teleconnections between the QBO influence and NH winter polar vortex (Holton-Tan effect), with clear impacts in early winter (November - January). A late winter reversal of this response (February - March) seen in the 1979 - 2016 analysis is not robust in the longer 1958 - 2016 period, highlighting the importance of using as long a data record as possible.

- There is no evidence for an early- or mid-winter QBO influence on SH vortex strength but good reanalysis agreement that the final SH warming occurs later in QBO-W than QBO-E when the phase is defined using 20 hPa QBO winds.

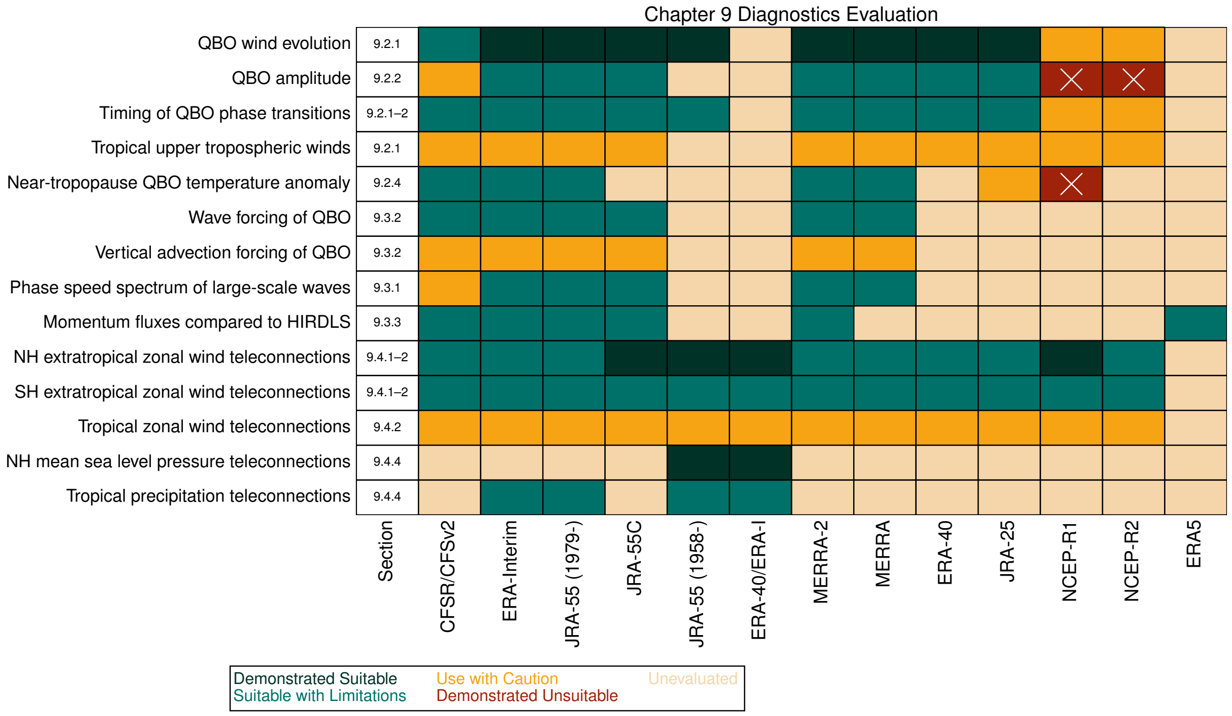
- In boreal winter there is a QBO impact on the strength of the tropical upper tropospheric winds of ~4 - 5 m s^{-1}, accompanied by an impact on the winter hemisphere subtropical jet near 30° latitude. There is good agreement of this signal for 1980 - 2016 in the four recent full-input reanalyses, but some details are not robust when the longer period 1958 - 2016 is examined.

- A QBO modulation of mean sea level pressure (MSLP) is found in NH winter over the extended 1958 - 2016 period in the JRA-55 reanalysis. The pattern, which in January resembles the North Atlantic Oscillation (NAO) pattern, is almost identical to that found in a recent study that combined ERA-40 / ERA-Interim to achieve a similarly long data record, suggesting that choosing either method for lengthening the data period is adequate for MSLP analysis.

- Analysis of the JRA-55 and ERA-Interim reanalyses over the satellite era demonstrate a QBO modulation of tropical precipitation, and both compare well with independent GPCC satellite observations. The response is mostly robust to inclusion of the pre-satellite years of JRA-55.

Recommendations from Chapter 9:

- Most reanalyses are suitable for determining the QBO phase but comparing several reanalyses is recommended for estimating the timing of phase transitions. MERRA-2 agrees best with the FUB at 30 hPa and is likely to provide the most accurate transition times at this level, but is a poor choice for 10 hPa QBO phase due to unusual features earlier in its record.



Chapter 9 Diagnostics Evaluation

	Section	CFSR/CFSv2	ERA-Interim	JRA-55 (1979-)	JRA-55C	JRA-55 (1958-)	ERA-40/ERA-I	MERRA-2	MERRA	ERA-40	JRA-25	NCEP-R1	NCEP-R2	ERA5
QBO wind evolution	9.2.1	🟩	⬛	⬛	⬛	⬛	⬜	⬛	⬛	⬛	⬛	🟧	🟧	⬜
QBO amplitude	9.2.2	🟧	🟩	🟩	🟩	⬜	⬜	🟩	🟩	🟩	🟩	❎	❎	⬜
Timing of QBO phase transitions	9.2.1–2	🟩	🟩	🟩	🟩	🟩	⬜	🟩	🟩	🟩	🟩	🟧	🟧	⬜
Tropical upper tropospheric winds	9.2.1	🟧	🟧	🟧	🟧	⬜	⬜	🟧	🟧	🟧	🟧	🟧	🟧	⬜
Near-tropopause QBO temperature anomaly	9.2.4	🟩	🟩	🟩	⬜	⬜	⬜	🟩	🟩	⬜	🟧	❎	⬜	⬜
Wave forcing of QBO	9.3.2	🟩	🟩	🟩	🟩	⬜	⬜	🟩	🟩	⬜	⬜	⬜	⬜	⬜
Vertical advection forcing of QBO	9.3.2	🟧	🟧	🟧	🟧	⬜	⬜	🟧	🟧	⬜	⬜	⬜	⬜	⬜
Phase speed spectrum of large-scale waves	9.3.1	🟧	🟩	🟩	🟩	⬜	⬜	🟩	🟩	⬜	⬜	⬜	⬜	⬜
Momentum fluxes compared to HIRDLS	9.3.3	🟩	🟩	🟩	🟩	⬜	⬜	🟩	⬜	⬜	⬜	⬜	⬜	🟩
NH extratropical zonal wind teleconnections	9.4.1–2	🟩	🟩	🟩	⬛	⬛	⬛	🟩	🟩	🟩	🟩	⬛	🟩	⬜
SH extratropical zonal wind teleconnections	9.4.1–2	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	⬜
Tropical zonal wind teleconnections	9.4.2	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	⬜
NH mean sea level pressure teleconnections	9.4.4	⬜	⬜	⬜	⬜	⬛	⬛	⬜	⬜	⬜	⬜	⬜	⬜	⬜
Tropical precipitation teleconnections	9.4.4	⬜	🟩	🟩	⬜	🟩	🟩	⬜	⬜	⬜	⬜	⬜	⬜	⬜

Demonstrated Suitable
Suitable with Limitations
Use with Caution
Demonstrated Unsuitable
Unevaluated

Figure 12.9: *(Same as **Figure 9.65**.) A summary of the diagnostics evaluated in Chapter 9: Quasi-Biennial Oscillation (QBO). The "Section" column at left indicates where in the chapter each diagnostic is described. See the beginning of this chapter for the meaning of the evaluation terms. "HIRDLS" = High Resolution Dynamics Limb Sounder (a satellite instrument); "NH" = Northern Hemisphere; "SH" = Southern Hemisphere.*

- The most recent reanalyses are recommended for comparison of QBO characteristics (amplitude, period, *etc.*) with climate models. JRA-55 provides the longest record and thus the most statistically robust estimates. MERRA-2 may also be a good choice because its representation of the QBO does not rely on the data assimilation to correct a severe model bias (*i.e.*, the lack of a QBO); however, at least based on the diagnostics presented here, this may not be important for most applications (and the aforementioned caveat about the 10-hPa level winds should also be noted). CFSR is less suitable than JRA-55, MERRA-2, or ERA-Interim because it underestimates the QBO amplitude compared to the other reanalyses.

- Conventional-input reanalyses are adequate for studies of the QBO as long as tropical radiosonde data are assimilated; JRA-55C appears to be as suitable for examining the QBO as JRA-55 (although its record is slightly shorter).

- Although not examined in the report, surface-input reanalyses (*e.g.,* ERA-20C) are not recommended for QBO studies. If a QBO exists in such a reanalysis it will be entirely produced by the forecast model; even if it is realistic, the lack of assimilated tropical stratospheric wind observations means that the QBO phase timing will almost certainly be incorrect.

- For studies of tropical temperature and meridional wind spectra any of the modern reanalyses are equally suitable since they show relatively small differences.

- For estimates of QBO wave forcing (*e.g.,* Eliassen-Palm flux divergence) care is required since there is substantial inter-reanalysis spread. Without suitable observations for validation it is not clear which reanalysis, if any, is most accurate, so comparison of several reanalyses is recommended. Given the very large natural (seasonal, inter-annual) variability of the QBO forcing terms, analysis of a long data period is recommended where appropriate.

- For QBO studies that involve the vertical advection term, comparison of as many of the modern reanalyses as possible is recommended because of large inter-reanalysis spread in vertical velocity in the lower tropical stratosphere. Model-level diagnostics are recommended since wave quantities can be damped by vertical interpolation. The post-1998 period is more reliable for evaluating wave spectra and QBO wave forcing.

- For investigation of QBO-vortex teleconnections we recommend using the longest available data records to maximise the signal-to-noise ratio (see also *Section 12.1.4*). However, while using pre-satellite era data to extend the data period is recommended for analysis of features at levels below ≈ 10 hPa, caution is required at the higher levels (*e.g.*, evaluating results from the pre- and post-satellite eras separately). For QBO studies of the SH, pre-satellite era data should be used with caution.

- For studies of the QBO impact on tropical / subtropical tropospheric circulation and surface precipitation the maximum available data period is recommended (*e.g.*, JRA-55 for 1958 - 2016 or concatenating the ERA-40 and ERA-Interim datasets). Care is required to distinguish the QBO signal from the ENSO signal.

- We recommend that reanalysis centres include 15 hPa and 40 hPa levels as standard output levels. The QBO amplitude peaks at 15 hPa in the FUB data, so model-reanalysis comparisons require this level for accurate validation of the models. The 40 hPa level, which is also in the FUB data, is highly correlated with the NH polar vortex response, and was the level at which the unusual easterly layer (the "QBO disruption") first emerged during 2015/16 NH winter.

12.1.8 Chapter 10: Polar Processes

In this chapter, we examined diagnostics of relevance to polar chemical processing and dynamics based on recent full-input reanalyses, including MERRA, MERRA-2, ERA-Interim, JRA-55, and CFSR/CFSv2. The selected diagnostics primarily target winter conditions. Observational datasets, reanalysis-driven CTM simulations, and operational analyses were also examined for some metrics. A summary evaluation of selected diagnostics examined in *Chapter 10* is provided in **Figure 12.10** as a quick reference to help users identify which reanalyses may be most suitable for a given issue related to stratospheric polar chemical processing. The key findings of this work, along with recommendations that follow from them, are summarized below.

Key findings of Chapter 10:

- In both polar regions, differences between temperatures from recent full-input reanalyses display an annual cycle. Using ERA-Interim as a reference, time series (2008 - 2013) of the differences in lower stratospheric daily polar-cap temperatures between the other reanalyses and the reference showed mainly positive deviations in summer but mainly negative deviations in winter, with the largest differences reaching ~ 1 K in the Antarctic and ~ 0.5 K in the Arctic. Thus, intercomparisons of the same reanalyses could find temperature discrepancies of opposite sign, depending on the season being examined.

- Polar winter temperatures from recent full-input reanalyses are in much better agreement in the lower and middle stratosphere than were those from older reanalysis systems.

- In the Southern Hemisphere especially, a dramatic convergence toward better agreement between the reanalyses is seen after 1999.

Average absolute differences from the reanalysis ensemble mean (REM) in wintertime daily minimum temperatures poleward of 40 ° S have been reduced from over 3 K prior to 1999 to generally less than 0.5 K in the most recent decade, while average differences in the area with temperatures below PSC thresholds have been reduced from over 1.5 % of a hemisphere to less than about 0.5 %. Other polar temperature and vortex diagnostics suggest a more complex picture, showing similar improvements for some reanalyses but persistent differences for others. The convergence toward better agreement is less apparent in the Northern Hemisphere.

- For many polar temperature and vortex diagnostics, reanalyses generally agree better in the Antarctic, where winters tend to have similar duration and potential for polar chemical processing every year, and thus the sensitivity to differences in meteorological conditions among reanalyses is low. In contrast, the generally warmer and more disturbed vortex and large interannual variability of Arctic winters lead to conditions that are frequently marginal, and thus the sensitivity to reanalysis differences is high.

- Comparisons of polar-cap averaged diabatic heating rates[4] in the lower stratosphere show that MERRA-2, ERA-Interim, JRA-55, and CFSR/CFSv2 give consistent results for the climatology and day-to-day evolution at pressures greater than about 20 hPa and should generally be suitable for polar processing studies.

- Comparisons of ERA-Interim, MERRA, and MERRA-2 with long-duration balloon observations in the Antarctic show that they reproduce the temperature and horizontal wind fluctuations of the balloons at about the 30 % level; thus a significant portion of the atmospheric gravity wave spectrum is not captured by the reanalyses.

- An evaluation of trajectory calculations from a Lagrangian transport model using long-duration balloon observations in the Antarctic found typical error growth rates of 60 - 170 km day^{-1} over 15-day trajectories for a subset of full-input reanalyses.

- Winter-long simulations from a chemistry transport model driven by different full-input reanalyses generally produce very similar results through most of the season for most species. However, substantial disparities between model runs are seen where composition gradients are largest. In particular, comparisons with satellite long-lived tracer measurements indicate that the model underestimates the strength of confined diabatic descent inside the winter polar vortex to varying degrees depending on the specific reanalysis used to force the model. As a consequence, considerable spread between the different simulations becomes evident by late winter.

- Estimates of chemical ozone loss based on satellite observations are relatively insensitive to the choice of reanalysis used to interpolate the measurements to isentropic surfaces and identify the vortex boundary. In contrast, chemical loss estimates based on simulated ozone fields from a chemistry transport model can differ substantially; a case study showed that forcing the model with different reanalyses yielded differences in the estimates of chemical ozone loss in the Antarctic vortex core as large as ~ 25 DU (20% - 30 %).

Recommendations from Chapter 10:

- Any of the recent full-input reanalyses (MERRA, MERRA-2, ERA-Interim, JRA-55, and CFSR/CFSv2) can be suitable for studies of lower stratospheric polar processing. However, substantial differences between the various reanalyses are found in some instances; therefore, the choice of which reanalysis to use in a given study may depend on the specific science questions being addressed.

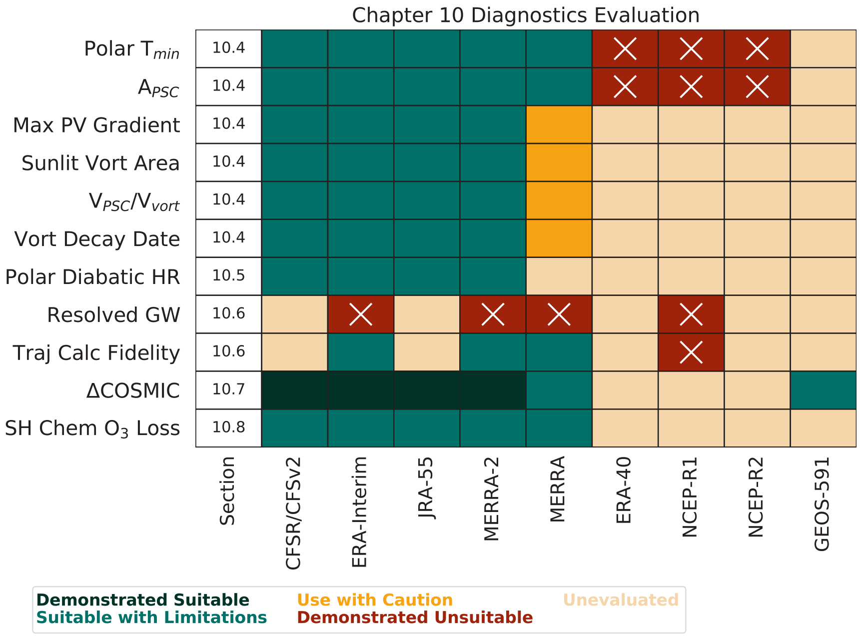
- Temperature biases in older meteorological reanalyses often rendered them unsuitable for accurately modeling interannual variability in PSC formation and consequent denitrification, chlorine activation, and chemical ozone loss; in particular, ERA-40, NCEP-NCAR R1, and NCEP-DOE R2 are obsolete and should no longer be used for studies of polar stratospheric chemical processing and dynamics.

- Because of the limitations of earlier reanalyses, it was not uncommon for modeling studies to try to match observed chlorine activation and/or ozone loss by imposing arbitrary systematic adjustments of 1 - 2 K or more on reanalysis temperatures. Increased confidence in the accuracy of current polar reanalysis temperatures provides tighter constraints on model parameterizations of microphysics/chemistry used to represent polar chemical processing. As a consequence, strong justification should be provided in modeling studies seeking to ascribe deficiencies in modeled chlorine activation and/or ozone loss to reanalysis temperature biases.

- Despite the overall good agreement between the polar temperatures from current full-input reanalyses, whenever feasible it is best to employ multiple reanalyses, even for studies involving recent winters for which differences between reanalyses are likely to be small; using more than one reanalysis allows estimation of uncertainties and the potential impact of those uncertainties on the results, especially for quantities that cannot be directly compared with observations.

---

[4] See the footnote on diabatic heating rates in reanalyses in *Section 12.1.3*.



	Section	CFSR/CFSv2	ERA-Interim	JRA-55	MERRA-2	MERRA	ERA-40	NCEP-R1	NCEP-R2	GEOS-5.9.1
Polar T_{min}	10.4	🟩	🟩	🟩	🟩	🟩	☒	☒	☒	⬜
A_{PSC}	10.4	🟩	🟩	🟩	🟩	🟩	☒	☒	☒	⬜
Max PV Gradient	10.4	🟩	🟩	🟩	🟩	🟧	⬜	⬜	⬜	⬜
Sunlit Vort Area	10.4	🟩	🟩	🟩	🟩	🟧	⬜	⬜	⬜	⬜
V_{PSC}/V_{vort}	10.4	🟩	🟩	🟩	🟩	🟧	⬜	⬜	⬜	⬜
Vort Decay Date	10.4	🟩	🟩	🟩	🟩	🟧	⬜	⬜	⬜	⬜
Polar Diabatic HR	10.5	🟩	🟩	🟩	🟩	⬜	⬜	⬜	⬜	⬜
Resolved GW	10.6	⬜	☒	⬜	☒	☒	⬜	☒	⬜	⬜
Traj Calc Fidelity	10.6	⬜	🟩	⬜	🟩	🟩	⬜	☒	⬜	⬜
ΔCOSMIC	10.7	■	■	■	■	🟩	⬜	⬜	⬜	🟩
SH Chem O_3 Loss	10.8	🟩	🟩	🟩	🟩	🟩	⬜	⬜	⬜	⬜

■ **Demonstrated Suitable** 🟩 **Suitable with Limitations** 🟧 **Use with Caution** ☒ **Demonstrated Unsuitable** ⬜ **Unevaluated**

Figure 12.10: *(Same as* ***Figure 10.26****.) A summary of the diagnostics evaluated in Chapter 10: Polar Processes. The "Section" column at left indicates where in the chapter each diagnostic is described. See the beginning of this chapter for the meaning of the evaluation terms. "Polar T_{min}" = minimum temperatures poleward of 40°; "A_{PSC}" = area of temperatures below PSC existence thresholds; "Max PV Gradient" = daily maximum gradients in potential vorticity, a measure of vortex strength; "Sunlit Vort Area" = area of the polar vortex in sunlight; "V_{PSC}/V_{vort}" = winter-mean volume of air with temperature below the nitric acid trihydrate PSC threshold, expressed as a fraction of the volume of air in the vortex; "Vort Decay Date" = the last day before which the vortex area is above 1 % of a hemisphere continuously for 30 days; "Polar Diabatic HR" = Diabatic heating rates in the polar vortex region; "Resolved GW" = resolved atmospheric gravity wave spectrum; "Traj Calc Fidelity" = fidelity of reanalysis-driven trajectory calculations from a Lagrangian transport model; "ΔCOSMIC" = differences between reanalysis and COSMIC GNSS-RO temperatures; "SH Chem O_3 Loss" = estimates of chemical loss in the Antarctic ozone hole from a chemistry transport model forced by reanalyses.*

- Reanalysis temperatures are generally unsuitable for assessment of trends in temperature-based diagnostics. Major changes in assimilated data inputs are often made at approximately the same time in all reanalyses, hindering determination of the impact of such changes through reanalysis intercomparisons. Caution is especially advised for the estimation of trends in diagnostics that aggregate low temperatures over months and/or vertical levels in the Northern Hemisphere, such as the winter-mean fraction of the vortex volume with air cold enough for PSCs to exist; such diagnostics are particularly sensitive to the specific PSC threshold chosen, which is subject to non-negligible interannual variability.

12.1.9 Chapter 11: Upper Stratosphere and Lower Mesosphere

In this chapter, we examined differences among reanalyses in the upper stratosphere and lower mesosphere among full-input reanalyses that provide data in this part of the atmosphere (MERRA, MERRA-2, ERA-40, ERA-Interim, JRA-25, JRA-55, and CFSR/CFSv2). A summary assessment of the diagnostics examined in this chapter is provided in **Figure 12.11**. Researchers interested in exploring a particular phenomenon within the USLM should consult the appropriate section of the chapter before proceeding, as indicated in the first column of this figure. Key findings and recommendations from this chapter are outlined below.

Key findings of Chapter 11:

- Differences among the reanalyses 1) decrease with time due to improvements in assimilated observational data, 2) increase with altitude due to differences in model top, sponge layers, and gravity wave drag treatments, and 3) increase nearer the Equator where sparse observations leave key dynamical phenomena largely unconstrained.

- Although no single reanalysis system is clearly better in representing all aspects of the USLM, higher-top systems such as MERRA and MERRA-2 are essential for capturing mesospheric circulation features such as the SAO and the QTDW.

- Differences in the satellite data assimilated into reanalyses as a function of time introduce discontinuities in both basic state variables and higher order diagnostics. This precludes trend studies based on a single reanalysis system.

- Differences in temperature among the reanalyses increase with height into the mesosphere at all latitudes. Likewise the inter-reanalysis differences in zonal wind increase with height especially in the equatorial region.

- Seasonal mean temperature differences defined with respect to MERRA are larger in older reanalyses (ERA-40 and JRA-25) and smaller in newer reanalyses (MERRA-2, ERA-Interim, and JRA-55).

- Westerly and easterly jets in the winter and summer stratosphere, respectively, are well reproduced in all of the evaluated reanalyses.

- The descending branch of the residual circulation in the winter stratosphere is strongest in MERRA, consistent with results prepared for *Chapter 5* (not shown; *Thomas Birner, personal communication*, 2021).

- Anomalous vertical temperature gradients around 3 hPa in JRA-25 lead to anomalous flow in the winter stratosphere. These features are not observed in the other reanalyses.

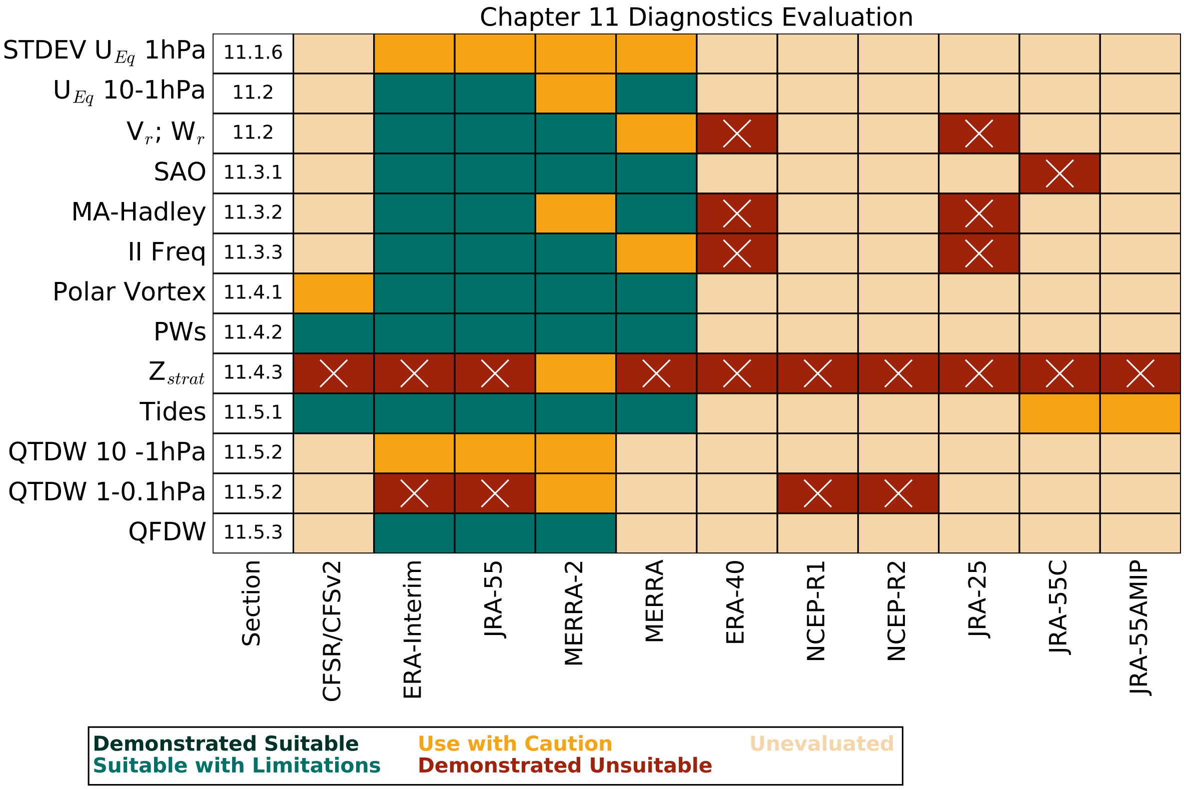
- Noisy meridional and vertical winds in ERA-40 cause larger dispersion of air parcels, which leads to "younger" age of air values and a weaker subtropical barrier in the stratosphere.

- Throughout the year, MERRA-2 has weaker cross-equatorial flow, a weaker middle-atmosphere Hadley circulation, and a westerly bias in the tropical USLM compared to ERA-Interim, JRA-55, and MERRA.

- Signatures of long-term variability due to the ENSO, the QBO, the 11-year solar cycle, and volcanic eruptions are evident in JRA-55, MERRA-2, and ERA-Interim; however, there are substantial differences among these reanalyses in the USLM, especially at equatorial latitudes.

- The mean SAO amplitude is reasonable in ERA-Interim, JRA-55, MERRA, and MERRA-2; comparison between JRA-55 and JRA-55C highlights the crucial role of assimilating satellite temperatures for accurately representing the SAO.

- The spatial patterns and magnitudes of inertial instability frequency are in good agreement among MERRA, MERRA-2, ERA-Interim, and JRA-55.



Chapter 11 Diagnostics Evaluation

	Section	CFSR/CFSv2	ERA-Interim	JRA-55	MERRA-2	MERRA	ERA-40	NCEP-R1	NCEP-R2	JRA-25	JRA-55C	JRA-55AMIP
STDEV U_{Eq} 1hPa	11.1.6	⬜	🟧	🟧	🟧	🟧	⬜	⬜	⬜	⬜	⬜	⬜
U_{Eq} 10-1hPa	11.2	⬜	🟩	🟩	🟧	🟩	⬜	⬜	⬜	⬜	⬜	⬜
V_r; W_r	11.2	⬜	🟩	🟩	🟩	🟧	☒	⬜	⬜	☒	⬜	⬜
SAO	11.3.1	⬜	🟩	🟩	🟩	🟩	⬜	⬜	⬜	⬜	☒	⬜
MA-Hadley	11.3.2	⬜	🟩	🟩	🟧	🟩	☒	⬜	⬜	☒	⬜	⬜
II Freq	11.3.3	⬜	🟩	🟩	🟩	🟧	☒	⬜	⬜	☒	⬜	⬜
Polar Vortex	11.4.1	🟧	🟩	🟩	🟩	🟩	⬜	⬜	⬜	⬜	⬜	⬜
PWs	11.4.2	🟩	🟩	🟩	🟩	🟩	⬜	⬜	⬜	⬜	⬜	⬜
Z_{strat}	11.4.3	☒	☒	☒	🟧	☒	☒	☒	☒	☒	☒	☒
Tides	11.5.1	🟩	🟩	🟩	🟩	🟩	⬜	⬜	⬜	⬜	🟧	🟧
QTDW 10 -1hPa	11.5.2	⬜	🟧	🟧	🟧	⬜	⬜	⬜	⬜	⬜	⬜	⬜
QTDW 1-0.1hPa	11.5.2	⬜	☒	☒	🟧	⬜	⬜	☒	☒	⬜	⬜	⬜
QFDW	11.5.3	⬜	🟩	🟩	🟩	⬜	⬜	⬜	⬜	⬜	⬜	⬜

Demonstrated Suitable **Use with Caution** 🟧 **Unevaluated** ⬜
Suitable with Limitations 🟩 **Demonstrated Unsuitable** ☒

Figure 12.11: *(Same as **Figure 11.52**.) A summary of the diagnostics evaluated in Chapter 11: Upper Stratosphere and Lower Mesosphere. The "Section" column at left indicates where in the chapter each diagnostic is described. See the beginning of this chapter for the meaning of the evaluation terms. Note that the score corresponding to "Demonstrated Suitable" was not assigned to any of the diagnostics listed here, so the darkest green colour does not appear in this table. The full names of the abbreviated diagnostics can be found in the Chapter 11 sections and subsections. Briefly, "STDEV" = the standard deviation; "U_{Eq}" = the zonal wind at the Equator; "V_r" and "W_r" = the residual circulation meridional and vertical velocities, respectively; "SAO" = the Semi-Annual Oscillation; "MA-Hadley" = the middle-atmosphere Hadley circulation; "II Freq" = the occurrence frequency of inertial instability; "PWs" = planetary waves; "Z_{strat}" = the height of the stratopause with emphasis on elevated stratopause events; "QTDW" = the quasi-2-day wave; "QFDW" = the quasi-5-day wave.*

- MERRA, MERRA-2, ERA-Interim, JRA-55, and CFSR/CFSv2 all capture multi-year winter mean polar vortex characteristics in both hemispheres, although CFSR wintertime vortex frequencies in the 50 ° to 70 ° latitude bands are 10 - 20 % lower in both hemispheres than those based on the other four reanalyses.

- MERRA, MERRA-2, ERA-Interim, JRA-55, and CFSR/CFSv2 sufficiently capture the multi-year mean seasonal evolution of the polar vortex at the stratopause during 2005 - 2015 (interannual variability is not assessed).

- Quasi-stationary PW-1 amplitudes show remarkable agreement among the reanalyses and with MLS observations in the extratropics during winter; larger differences are seen at lower latitudes and during summer.

- Elevated Stratopause (ES) events represent strong, transient departures from climatological conditions in the Arctic USLM. These events are generally unconstrained by observations (with the exception of MERRA-2 which assimilates temperatures from Aura MLS after 2004). Their representation in all reanalyses depends strongly on the nature of the sponge layer in the forecast model used to produce the reanalysis, and thus cannot be regarded as trustworthy.

- While reanalyses reproduce the global patterns of the diurnal and semi-diurnal migrating tides, their amplitudes are underestimated by 20 - 50 % compared to SABER observations.

- The representation of the quasi-2-day wave is qualitatively similar in MERRA-2, ERA-Interim, and JRA-55, but with 50 % differences in amplitude.

- There is excellent agreement in the representation of the quasi-5-day wave among MERRA-2, ERA-Interim, and JRA-55, suggesting that the origin and propagation of this wave involve stratospheric processes that are well represented in these systems.

Recommendations from Chapter 11:

- Scientific studies using reanalyses in the USLM should make every effort to also include comparisons with independent observations. This imperative will require sustained engagement from reanalysis data users, new observational campaigns and operational measurement platforms for evaluation of reanalysis data, and a renewed commitment to replace the aging satellites currently relied upon for temperature and constituent observations of the middle atmosphere.

- Large discontinuities that occur due to differences in the data assimilation process preclude trend studies based on any single reanalysis system.

- There are large temperature and wind differences among the reanalyses in the tropical USLM. Using two or more reanalyses datasets to study phenomena (*e.g.*, the SAO, the diurnal tide) in this region of the atmosphere is recommended to increase confidence.

- There are large uncertainties in MERRA-2 zonal winds in the tropics; MERRA, ERA-Interim, and JRA-55 are in better agreement with each other up to 1 hPa than they are with MERRA-2.

- There are large uncertainties in “older” reanalysis datasets in the USLM; the meridional circulation in the stratosphere and mesosphere is more realistic in MERRA-2, ERA-Interim, and JRA-55 than in MERRA, ERA-40, and JRA-25.

- Both Eulerian-mean and residual-mean meridional flows in ERA-40 are noisier than those in the other reanalyses, thus, science studies based on ERA-40 residual circulation velocities would likely generate noisier results.

- JRA-55C is not suitable for studies of the SAO.

- Low polar vortex frequency of occurrence biases in CFSR/CFSv2, due to high polar temperatures and a weak polar night jet, render this reanalysis dataset less suitable for polar vortex studies compared to MERRA, MERRA-2, ERA-Interim, or JRA-55.

- MERRA, MERRA-2, ERA-Interim, JRA-55, and CFSR/CFSv2 are all suitable for studying quasi-stationary PW-1 patterns in the winter extratropics, but care should be exercised for studies focusing on the subtropics or the summer.

- Reanalyses should not be relied upon for studying ES events. Even for MERRA-2, the underlying forecast model does not capture the evolution of ES events correctly and so derived quantities (other than temperatures that are directly assimilated) should be treated with caution.

- Older reanalyses such as ERA-40 or JRA-25 are not suitable for tidal studies.

- Tidal results should not be extrapolated from one time to another as the representation of tides is sensitive to the satellite data assimilation.

- There are large uncertainties in using reanalysis data to study 5-day and 2-day wave normal modes; different reanalyses may yield different results.

12.2 Overall findings and reanalysis user recommendations

Several common findings and recommendations emerge from the detailed and extensive reanalysis comparisons described in *Chapters 3–11* and summarized above:

- All studies find substantial improvements in the most recent generation of reanalyses, even in cases where the older reanalyses are adequate for some diagnostics. We thus recommend that studies using full-input reanalyses be done with CFSR/CFSv2, ERA-Interim (and/or ERA5), JRA-55, and/or MERRA-2 rather than reanalyses from previous generations.

- In particular, NCEP-NCAR R1 and NCEP-DOE R2 are inadequate for many diagnostics. These reanalyses are deprecated based not only on the findings of this activity, but also on a wealth of comparisons spanning more than a decade. With the availability of modern reanalyses providing coverage of the pre-satellite (*e.g.*, JRA-55 and ERA5) and the ability to obtain those reanalyses on coarser grids, there should be no reason to continue using these older reanalyses.

- A number of studies find deficiencies in CFSR/CFSv2 relative to its peers (ERA-Interim, JRA-55, MERRA-2), and in some cases the changes between CFSR and CFSv2 are sufficiently large that it may not be appropriate to use these two datasets as if they were continuous.

- Several studies have shown that valuable information on the pre-satellite era can be obtained from conventional-input reanalyses that do not assimilate satellite data. For such studies, it is essential to first assess how the diagnostics in question compare with the full-input reanalysis from the same system during the satellite era.

- The vast majority of studies described herein found scientific benefit in using multiple reanalyses and comparing the results. This type of approach is especially important for diagnostics that cannot be directly compared with observations. In cases where the reanalyses agree well, results based on multiple reanalyses still provide valuable uncertainty estimates.

- All reanalyses show some level of discontinuities related to major changes in data inputs, a key example being changes associated with (and improvements in reanalysis agreement after) the switch from TOVS to ATOVS around 1998/1999. As different assimilation systems handle these changes in different ways, the impacts also differ across reanalyses.

- While several studies reported herein show valuable information obtained from studying trends in reanalysis data, great caution should be used in conducting such studies, not least because of the impacts of observing system changes as mentioned in the previous point. Trend studies should always compare multiple reanalyses, and consistency among results for multiple reanalyses should be viewed as a necessary but not sufficient condition for robustness.

- Many of the studies described herein benefitted from using the highest vertical resolution available and, for some (especially studies of conditions at and around the tropopause), this high vertical resolution proved critical. We thus recommend using reanalysis products on model levels in all analyses for which sharp vertical gradients or fine-scale vertical features may be important.

- Several quantities (notably diabatic heating rates[5], ozone and water vapour, and products related to clouds and convection) are handled and reported very differently across different reanalyses. Careful consideration of how the individual products are produced is necessary when using and comparing them.

- Several chapters have emphasized the importance of continuing data records (especially satellite trace gas data), for which ongoing records are in jeopardy due to aging instruments and an uncertain commitment to future missions. These data are essential benchmarks for evaluating reanalysis data both directly (validation of reanalysis products) and indirectly (evaluation of reanalysis-driven or nudged CTM and CCM simulations). In addition, several currently available homogenized satellite datasets have been shown to provide important improvements in reanalysis products when assimilated, so continuing (and improving upon) records such as these should be a priority.

12.3 Recommendations for improving reanalyses and their evaluation

One important aspect of the S-RIP activity was the involvement of reanalysis centres, as well as the continuing dialog between representatives of these centres and the reanalysis data users who conducted studies for S-RIP. A number of recommendations for future work have emerged from these interactions, including recommendations related to future reanalysis development, improvements in the output products, data formats, or grids, and the need for further observations both for assimilation into reanalysis systems and for evaluation of reanalysis products.

---

[5] See the footnote on diabatic heating rates in reanalyses in *Section 12.1.3*.

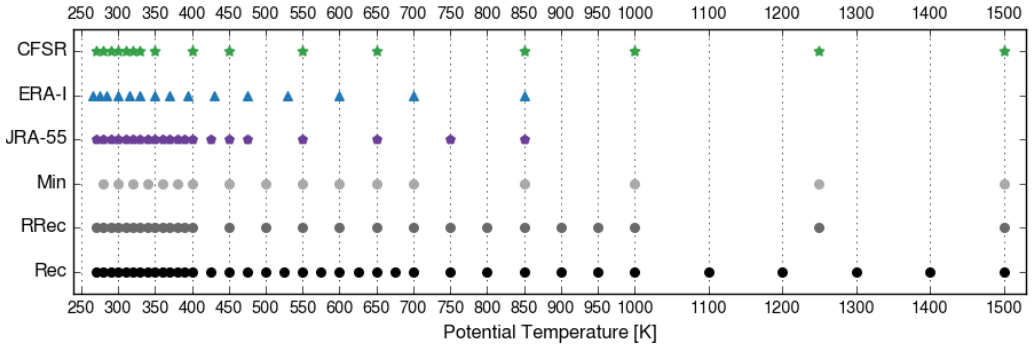


Figure 12.12: Current and proposed isentropic levels from the surface through the midstratosphere. The following three sets were proposed for the survey (indicates levels that are above the top of some of the most recent models/analyses).* ***Min:*** *280, 300, 320, 340, 360, 380, 400, 450, 500, 550, 600, 650, 700, 850, 1000, 1250, 1500, 1750, 2000, 2500, 3000, 3500*, 4000* K.* ***RRec:*** *270, 280, 290, 300, 310, 320, 330, 340, 350, 360, 370, 380, 390, 400, 450, 500, 550, 600, 650, 700, 750, 800 ,850, 900, 950, 1000, 1250, 1500, 1750, 2000, 2500, 3000, 3500*, 4000* K.* ***Rec:*** *270, 280, 290, 300, 310, 320, 330, 340, 350, 360, 370, 380, 390, 400, 425, 450, 475, 500, 525, 550, 575, 600, 625, 650, 675, 700, 750, 800, 850, 900, 950, 1000, 1100, 1200, 1300, 1400, 1500, 1750, 2000, 2250, 2500, 3000, 3500*, 4000* K. Based on the results of this survey, we recommend RRec (the middle resolution).*

12.3.1 S-RIP survey results and related product needs from S-RIP studies

To help clarify the output product format needed by reanalysis users, we conducted several surveys related to the adequacy of currently available products and output grids. The results of these surveys are summarized briefly below. The number of respondents was 28. Overall, 63 % of respondents expect to need reanalysis data at or near the resolution of the native model grids, while 74 % of respondents need data either on model levels or on isobaric or isentropic grids that are finer or more extensive than currently available. Detailed surveys on user needs for data on isentropic and isobaric levels resulted in the following:

Isentropic Levels:

- Approximately 70 % of respondents need data on isentropic surfaces.
- Of those, 82 % say the currently available levels are inadequate.
- Of the three sets of levels we proposed (**Figures 12.12** and **12.13**), 28 % say they need the finest resolution, 44 % the medium resolution, and 28 % the coarsest resolution.
- **Products most needed on isentropic surfaces:**
  - Pressure / Temperature: 100 %
  - Potential Vorticity: 94 %
  - Zonal and Meridional Winds: 89 %
  - Ozone mixing ratio: 66 %
  - Specific Humidity: 50 %
  - Montgomery Streamfunction: 44 %

The grey dots in **Figures 12.12** and **12.13** show the three sets of common isentropic levels we proposed for the survey. Based on the survey results, among Min, RRec, and Rec (see **Figures 12.12** and **12.13** for their definitions), we recommend the following set RRec:

RRec: 270, 280, 290, 300, 310, 320, 330, 340, 350, 360, 370, 380, 390, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1250, 1500, 1750, 2000, 2500, 3000, 3500*, 4000* K

where * indicates levels that are above the top of some of the most recent models/analyses.

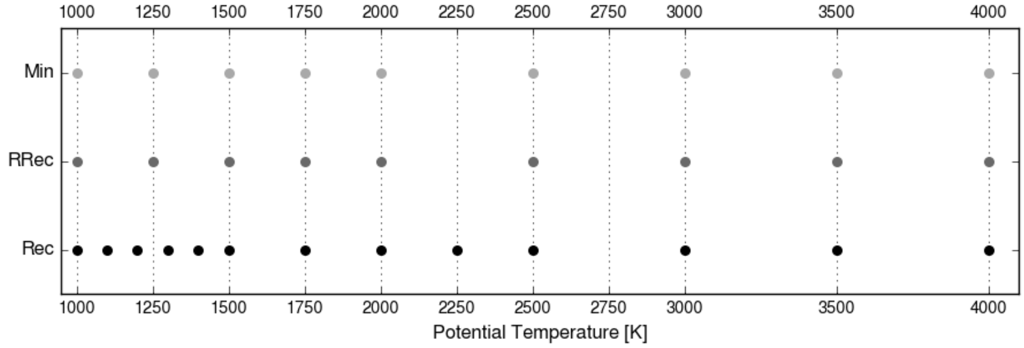


Figure 12.13: As for ***Figure 12.12****, but for proposed isentropic levels in the USLM. Based on the results of our survey, we recommend RRec (the middle resolution).*

Pressure Levels:

- Approximately 81 % of respondents need data on pressure surfaces.
- Of those, 84 % say the currently available levels are inadequate.
- 95 % say the proposed additional levels (see below) would be useful to them.

The standard ERA-Interim output diagnostic levels are:

1000, 975, 950, 925, 900, 875, 850, 825, 800, 775, 750, 700, 650, 600, 550, 500, 450, 400, 350, 300, 250, 225, 200, 175, 150, 125, 100, 70, 50, 30, 20, 10, 7, 5, 3, 2, 1 hPa

Other recent full-input reanalyses use very similar pressure-level grids for data distribution.

We propose additional levels at 85, 60, 40, and 15 hPa (to improve resolution in the vicinity of the tropical tropopause and the QBO) and 0.7, 0.3, 0.1, 0.03, and 0.01 hPa (to improve coverage of the USLM).

In summary, our recommendation on common pressure levels for future reanalysis products is as follows:

1000, 975, 950, 925, 900, 875, 850, 825, 800, 775, 750, 700, 650, 600, 550, 500, 450, 400, 350, 300, 250, 225, 200, 175, 150, 125, 100, 85, 70, 60, 50, 40, 30, 20, 15, 10, 7, 5, 3, 2, 1, 0.7, 0.3, 0.1, 0.03, 0.01 hPa

Figure 12.14 illustrates the vertical grid spacing for these requested pressure levels, as well as vertical grid spacings for the current standard pressure levels and the model levels used in producing ERA-Interim and ERA5. Other levels suggested by survey respondents but not represented in our recommendation are 80 hPa for UTLS studies and 0.5 hPa and 0.2 hPa for USLM studies.

In addition, both the surveys conducted and many of the S-RIP studies reported herein suggest community needs for:

- Diabatic heating rates from all physics on model grids, with all reanalyses reporting a consistent minimum product. Currently some reanalyses report only LW and SW heating rates on model levels, whereas others also report heating rates from all physics. There are also other differences in how the reanalysis provide diabatic heating rates that make them difficult to use and compare. Diabatic heating rates are critical for transport studies, especially in the upper troposphere and stratosphere. For consistency and comparability, it would be helpful for future reanalyses to provide (on model levels) temperature tendencies from (1) all physics, (2) all-sky radiation, and (3) clear-sky radiation, with the latter two provided separately for the LW and SW components. Additional terms (*e.g.*, convection, large-scale condensation, turbulence, assimilation, *etc.*) are also valuable for evaluating individual reanalyses and conducting scientific studies, and we suggest that these terms be provided as computational resources and model formulation permit.
- Integration with satellite simulators where possible. The inclusion of the Cloud Feedback Model Intercomparison Project (CFMIP) Observation Simulator Package (COSP) in MERRA-2 provided valuable context not only for observational validation, but also for understanding differences between MERRA-2 and other reanalysis cloud products. As these simulators and their use in climate model evaluation expands, their application to reanalysis products becomes increasingly relevant. The provision of model-resolution reanalysis outputs at high frequency is a welcome step toward facilitating offline application of satellite simulators. However, computational resources permitting, full integration within the reanalysis model would go a long way toward enabling wider and more effective use of these tools.

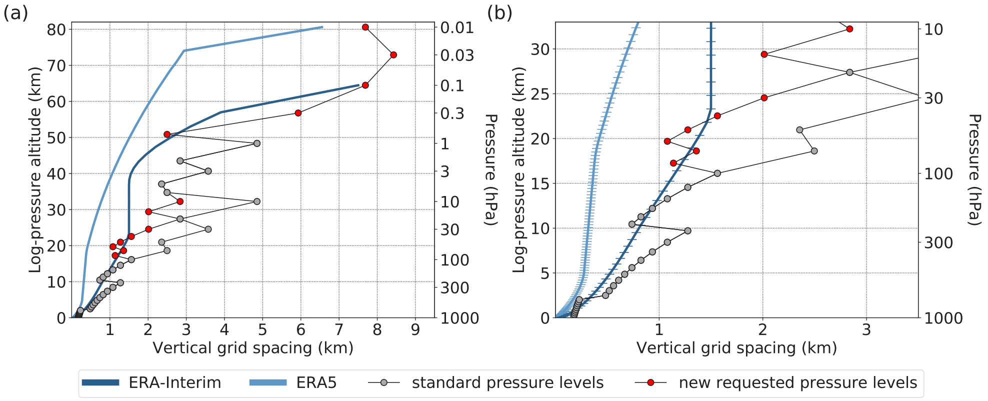


Figure 12.14: *Vertical profiles of vertical grid spacing for the requested pressure levels (see text), as well as vertical grid spacings for current standard pressure levels (from ERA-Interim; see text) and model levels (from ERA-Interim and ERA5). Panel (a) shows vertical grid spacing from the surface to 0.01 hPa (illustrating the proposed extension of the vertical grid), while panel (b) provides a zoomed view from the surface to 10 hPa (illustrating the requested finer resolution around the tropical tropopause and the lower part of the QBO).*

- Information on uncertainty estimates for reanalysis products (especially basic fields such as temperatures and winds). Despite recent advances in this regard (*e.g.*, the ensemble of data assimilations produced as part of ERA5), such estimates remain problematic to produce for complex data assimilation systems. It may therefore be a useful goal for continuing S-RIP efforts to produce such estimates based on intercomparisons.

- Availability of a common data format for all reanalyses. This would be in line with current practice in the climate modelling community (*e.g.*, CMIP), for which a common data format has proved invaluable for intercomparison studies. Adoption of community standards for variable and file metadata in tools for preprocessing reanalysis data prior to download (see below) would boost the effectiveness of these tools and support the evaluation and intercomparison of future reanalysis products.

12.3.2 Data access issues

Ease of access to reanalysis datasets on multiple grids has improved greatly over the years of the S-RIP activity. We note and commend recent efforts to improve accessibility to both model-level and native-grid products for studies where resolution is critical and reduced-resolution products for cases where resolution is not critical and disk space or bandwidth are limiting factors. Nevertheless, the ever-increasing data volume for new reanalyses remains the largest current and future challenge to data access, as illustrated by difficulties in obtaining, storing, and processing ERA5 data at high resolution and on model levels. It will be essential to devise solutions for these challenges, not least in light of the numerous cases documented in this report for which the high resolution that engenders such large file sizes proved important both for fair evaluation of the reanalyses and for clarifying understanding of the diagnostics under evaluation.

We envision solutions for this issue taking multiple forms, from simple improvements in procedures and infrastructure to extensive investment in distributed processing and server-side applications. Developments in the latter direction have been extremely valuable for the S-RIP activity, as some reanalysis centres have begun dedicating computational resources for users to conduct simple preprocessing steps (*e.g.*, regridding, subsampling, and temporal averaging) prior to downloading data. These tools reduce the computational overhead for reanalysis data users, thus speeding up analysis and allowing access to a more complete set of reanalysis products for cross-validation and hypothesis testing. Particularly useful features include options for remapping data onto user-selected grids, subsetting regions or variables, and daily averaging. We express here our appreciation for the resources and hard work that reanalysis centres and their employees have put into making these tools available, as well as our wholehearted support for further investment in this direction.

Many participants of S-RIP have also benefitted from a designated group workspace on the JASMIN “super-data-cluster” in the United Kingdom. Funded by the National Environment Research Council (NERC) and the UK Space Agency, this platform provided data storage and analysis tools that facilitated some of the more computationally intensive tasks undertaken by S-RIP participants. Resources permitting, further investment in the server-side tools provided by reanalysis centres might adopt some of the capabilities of this type of group workspace, such as temporary storage of intermediate products and/or more flexible pre-processing tools. Such developments would be invaluable for making new reanalyses accessible to a wider community of data users.

Helpful steps for improving data access can also be taken without requiring large investments of funds or computational resources. For example, standard sets of metadata to support widely used scripting tools (*e.g.*, parameter tables or grid description files for use with the Climate Data Operators developed at the Max-Planck-Institut für Meteorologie) could be used to construct ‘recipes’ for users to convert data from a more concise, centre-preferred format (*e.g.*, GRIB) to a more verbose and user-friendly format (*e.g.*, CF-compliant NetCDF4 with standard naming conventions). Such recipes could be organized by and provided together with pre-defined data collections (*e.g.*, upper-air analysis, forecast diagnostics, *etc*.), as was previously done for some reanalyses using GrADS control files. Regular testing of output data against common data processing tools and manipulations (*e.g.*, remapping, area selection, merging or averaging in time) would also be helpful. Often small adjustments to the grid description or other aspects of the file metadata are all that is needed to ensure compatibility with a wide range of software tools for climate data analysis. Although these steps cannot address barriers associated with computational overhead, they can substantially reduce the ‘learning curve’ for users interested in adopting and applying a new reanalysis dataset. Community resources like S-RIP (presuming it continues in some form) and reanalyses.org can also play valuable roles in creating, updating, and distributing these types of tools.

12.3.3 Documentation issues

It is critical for information on the models and assimilation systems to be kept current and accessible. In the past, documentation for reanalyses has often been sparse, out-of-date, difficult to find, or all of the above. For some centres (notably ECMWF), this situation has improved in recent years. It is important to have information available both on the ideas and assumptions behind the original model schemes (generally accessible in some form now, though not always easy to find), and on how those schemes have evolved since their original publication, in some cases 20 - 30 years ago (generally not available now).

We hope that one legacy of S-RIP will be to provide a model for immediately, consistently, and systematically documenting each new reanalysis, and for bringing and keeping documentation on existing reanalyses up to date. The detailed information presented in *Chapter 2: Description of the Reanalysis Systems* could serve as a template in this regard.

12.4 Prospects for the future

S-RIP was originally planned to continue until 2018 (*i.e.*, 5 years starting from the Planning Meeting in 2013). However, a fundamental goal of S-RIP is to provide well-organized feedback to the reanalysis centres, thus forming a "virtuous circle" of assessment, improvements in reanalyses, further assessment, and further improvements in reanalyses. To this end, calculations of diagnostics suited to numerous types of studies have been and are being developed for current reanalyses. These diagnostics can then be easily extended and applied to the assessment of future reanalyses. Since most reanalysis centres have ongoing programmes to deliver new and improved reanalyses, it may be valuable to continue S-RIP beyond this initial period of 8 years. The SPARC SSG meeting in 2022 will therefore provide a critical opportunity for that body to review the value of S-RIP, with input from the reanalysis centres and atmospheric science and climate researchers, and discuss how the continuing goals of systematic evaluation of reanalyses can be supported into the future.

Regardless of the future development of S-RIP, it is important for this project to leave a lasting legacy through publication of its report that helps to sustain international interest in the assessment of reanalyses. A primary goal of the project is to establish tighter links between reanalysis providers and SPARC-related researchers. It is thus hoped that outcomes from the S-RIP assessment will facilitate and even drive future reanalysis developments in a systematic, standardised way, in place of the ad hoc approaches that have been used previously. A further legacy will be the creation of public archives (at BADC/CEDA and NOAA; see *Chapter 1*, *Section 1.5*) of processed reanalysis data with standard formats and resolutions, which will help to enable both further intercomparisons and scientific analyses without repetition of expensive pre-processing steps. This ensemble of derived data sets is freely available to researchers worldwide, and is intended to be a useful tool for reanalyses assessment beyond the lifetime of the project.

Acknowledgements

We acknowledge the scientific guidance and sponsorship of the World Climate Research Programme, coordinated in the framework of SPARC. We are grateful to Theodore Shepherd and Greg Bodeker (past SPARC co-chairs) for strong encouragement and valuable advice during the proposal phase of the project during 2011 - 2012, and Joan Alexander (serving SPARC co-chair during 2013 - 2015), Neil Harris (serving SPARC co-chair since 2014), Judith Perlwitz (serving SPARC co-chair during 2016 - 2020), and Seok-Woo Son (serving SPARC co-chair since 2019) for continued support and encouragement. We are also very grateful to the review handling editors of this report, Judith Perlwitz, Seok-Woo Son, Vincent-Henri Peuch, Karen Rosenlof, and Gufran Beig, and to the reviewers. The Information Initiative Center of Hokkaido University, Japan, has hosted the S-RIP web server since 2014. We thank the reanalysis centres for providing their support and data products. The British Atmospheric Data Centre (BADC) of the UK Centre for Environmental Data Analysis (CEDA) has provided a virtual machine for data processing, a group workspace for data storage, and public data archive sites. We thank Yulia Zyulyaeva for contributions to S-RIP as an original member of the working group, as a chapter colead through October 2014, and as the designer of the S-RIP logo. We thank Diane Pendlebury for contributions to S-RIP as a chapter co-lead through August 2015. We thank Quentin Errera, the lead of the SPARC Data Assimilation working group, for co-organizing the 2014, 2015, 2016, and 2017 S-RIP workshops. We also thank John McCormack for co-organizing the 2017 workshop. Travel support for some participants of the 2013 planning meeting, the 2014, 2015, 2016, and 2017 workshops, and the 2018 chapter-lead meeting was provided by SPARC. We thank Peter Haynes and Gabriele Stiller for their work as guest editors for the special issue "The SPARC Reanalysis Intercomparison Project (S-RIP)" in the journals Atmospheric Chemistry and Physics (ACP) and Earth System Science Data (ESSD); we offer special thanks to William Lahoz (1960 - 2019) for his role as the other guest editor of this special issue, which he performed excellently even as he battled his final illness. We thank the hard works at the SPARC Office, in particular Mareike Heckl, Brigitte Ziegele, and Sabrina Zechlau for their typesetting of the report. Work at the Jet Propulsion Laboratory, California Institute of Technology, was carried out under a contract with the National Aeronautics and Space Administration.

Major abbreviations and terms

20CR	20th Century Reanalysis of NOAA and CIRES
3D	three dimensional
AIRS	Atmospheric Infrared Sounder
Alt	altitude
AMSU	Advanced Microwave Sounding Unit
AoA	Age of Air
A_{PSC}	area of temperatures below PSC existence thresholds
ATMS	Advanced Technology Microwave Sounder
ATOVS	Advanced TIROS Operational Vertical Sounder
BADC	British Atmospheric Data Centre
BDC	Brewer-Dobson Circulation
Calc	Calculation
CCM	Chemistry-Climate Model
CDR	Climate Data Record
CEDA	Centre for Environmental Data Analysis
CF	Climate and Forecast
CFMIP	Cloud Feedback Model Intercomparison Project
CFSR	Climate Forecast System Reanalysis of NCEP
CFSv2	Climate Forecast System, version 2
Ch	Channel (e.g., Ch1: Channel 1)
Chem	Chemical
CIRES	Cooperative Institute for Research in Environmental Sciences (NOAA and University of Colorado Boulder)
Clim	Climatology
CMIP	Coupled Model Intercomparison Project
COSMIC	Constellation Observing System for Meteorology Ionosphere and Climate
COSP	CFMIP Observation Simulator Package
CP	Cold Point
CPT	Cold-Point Tropopause
CTM	Chemistry-Transport Model
CWC	Cloud Water Content
DOE	Department of Energy
DU	Dobson unit
Dyn	Dynamical
ECMWF	European Centre for Medium-Range Weather Forecasts
ENSO	El Niño–Southern Oscillation
E-P flux	Eliassen-Palm flux
EqL	Equivalent Latitude
ERA-20C	ECMWF 20th century reanalysis
ERA-40	ECMWF 40-year reanalysis
ERA5	the fifth major global reanalysis produced by ECMWF
ERA-Interim (or ERA-I)	ECMWF interim reanalysis
ES	Elevated Stratopause
ExUTLS	Extratropical Upper Troposphere and Lower Stratosphere
FLXHR	Level 2B Fluxes and Heating Rates of CloudSat data product

FUB	Freie Universität Berlin
GEOS-5	Goddard Earth Observing System Model of the NASA, Version 5
GNSS-RO	Global Navigation Satellite System Radio Occultation
GPCC	Global Precipitation Climatology Centre
GrADS	Grid Analysis and Display System
GRIB	GRIdded Binary or General Regularly-distributed Information in Binary form
HCC	High Cloud Cover
HIRDLS	High Resolution Dynamics Limb Sounder
IAU	Incremental Analysis Update
ITCZ	Intertropical Convergence Zone
JASMIN	a data intensive supercomputer for environmental science at United Kingdom
Jet Rel	in coordinates relative to the subtropical jet core location
JRA-25	Japanese 25-year Reanalysis
JRA-55	Japanese 55-year Reanalysis
JRA-55AMIP	Japanese 55-year Reanalysis based on AMIP-type simulations
JRA-55C	Japanese 55-year Reanalysis assimilating Conventional observations only
LRT	Lapse-Rate Tropopause
LS	Lower Stratosphere
LW	Long-Wave
LWCRE	Long-Wave Cloud Radiative Effect
LZRH	Level of Zero net Radiative Heating
MA-Hadley	the middle-atmosphere Hadley circulation
MERRA	Modern Era Retrospective-Analysis for Research and Applications
MERRA-2	Modern Era Retrospective-Analysis for Research and Applications, Version 2
MERRA-2-ANA	MERRA-2 “analysis” data products that result directly from the Gridpoint Statistical Interpolation (GSI) analyses
MERRA-2-ASM	MERRA-2 “assimilation” data products that are the result of applying the Incremental Analysis Update (IAU)
MIPAS	Michelson Interferometer for Passive Atmospheric Sounding
MLS	Microwave Limb Sounder
MSLP	Mean Sea Level Pressure
MSU	Microwave Sounding Unit
MTp	Multiple Tropopause
NAO	North Atlantic Oscillation
NASA	National Aeronautics and Space Administration
NCAR	National Center for Atmospheric Research
NCEP	National Centers for Environmental Prediction of the NOAA
NERC	National Environment Research Council of the United Kingdom
NetCDF4	Network Common Data Form, Version 4
NH	Northern Hemisphere
NOAA	National Oceanic and Atmospheric Administration
OLR	Outgoing Longwave Radiation
P	pressure
PSC	Polar Stratospheric Cloud
PV	Potential Vorticity
PW-1	Planetary Wave-1 (wavenumber one of planetary waves)
PWs	Planetary Waves
QBO	Quasi-Biennial Oscillation
QBO-E	QBO Easterly phase
QBO-W	QBO Westerly phase
QFDW	Quasi 5-Day Wave
QTDW	Quasi 2-Day Wave
R1 (or NCEP-R1)	NCEP-NCAR Reanalysis 1

R2 (or NCEP-R2)	NCEP-DOE Reanalysis 2
RCTT	Residual Circulation Transit Time
Rel	Relative
REM	Reanalysis Ensemble Mean
RO	Radio Occultation
RRec	the final recommended set of isentropic levels
SABER	Sounding of the Atmosphere using Broadband Emission Radiometry
SAO	Semi-Annual Oscillation
SASM	South Asian Summer Monsoon
SH	Southern Hemisphere
SPARC	Stratosphere-troposphere Processes And their Role in Climate
S-RIP	SPARC Reanalysis Intercomparison Project
SSG	Scientific Steering Group
SSU	Stratospheric Sounding Unit
SSW	Sudden Stratospheric Warmings
STDEV	Standard Deviation
STE	Stratosphere-Troposphere Exchange
SubV	subvortex
SW	Short-Wave
SWV	Stratospheric Water Vapour
TCO	Total Column Ozone
TIROS	Television Infrared Observation Satellite
T_{min}	minimum temperatures
TOVS	TIROS Operational Vertical Sounder
T	temperature
Tp	Tropopause
Traj	Trajectory
TTL	Tropical Tropopause Layer
U	zonal wind
U_{Eq}	the zonal wind at the Equator
USLM	Upper Stratosphere and Lower Mesosphere
UT	upper troposphere / upper tropospheric
UTLS	Upper Troposphere and Lower Stratosphere
V_{PSC}	winter-mean volume of air with temperature below the nitric acid trihydrate PSC threshold
V_{vort}	volume of air in the vortex
V_r	the residual circulation meridional velocity
W_r	the residual circulation vertical velocity
WV	water vapour
Yr	year
ZM	Zonal Mean
Z_{strat}	the height of the stratopause with emphasis on elevated stratopause events