Dear Prof. Haynes,

We've made a number of changes to our manuscript in response to the two reviews, and feedback that we received from other members in the SPARC Reanalysis Intercomparison Project. We respond to the reviews in detail below, but wanted to highlight a few substantive changes first.

(1) It was brought to our attention that NCEP made very large changes to their reanalysis system, which came online on 1 January 2011. It is best practice to consider their reanalysis from 1979 to 2010 (NCEP CFSR) as distinct from their renalaysis from 2011 onwards, which is called NCEP CFSv2. We checked and found that this did not substantially impact the annular modes, though there is improved coherence between CFSv2 with the other modern reanalyses as compared to CFSR, which is now highlighted in a new Figure 5. To emphasize this difference between CFSR and CFSv2 to the community, we adjusted the analysis periods for Figs. 3 and 4 (previously figures 2-3).
(2) ERA5 output has become available for a limited period. We know include some preliminary assessment of this reanalysis.
(3) It was suggested that we could compare the Southern Annular Mode time series with a station-based index established by Gareth Marshall. We can only look at the monthly mean near the surface, but it did reveal that the surface-input reanalyses are superior to the full input reanalyses in the pre-satellite era. This is included in section 6 and featured in a new table 2.
(4) Both of the reviewers were concerned that our polar-cap average annular mode definition would only track variability over the pole. To clarify that it captures the full structure of the annular mode, we illustrate the associated patterns on a few select levels for two reanalyses in the new Figure 2

We thank both reviewers for their feedback on the manuscript, and believe it has improved in response to their feedback. We hope that the study is now ready for publication in ACP.

Sincerely,

Edwin Gerber and Patrick Martineau.

Detailed response to reviewer 1.

Thank you for these detailed comments our manuscript. We respond below to your comments of a scientific nature, and have corrected the grammatical and typographical mistakes from the previous draft.

In the context of troposphere-stratosphere coupling, it would have been useful to see more Southern hemisphere analysis, particularly associated with the final warming.

We appreciate the author's concern about neglecting austral hemisphere. We avoided the final warming in part because several authors have noted that the response to the Stratospheric Final Warming (SFW), is structurally different from the annular modes (Black and McDaniel 2007a,b and Sheshadri et al. 2013). In addition, the SFW is being explored as part of SPARC Reanalysis Intercomparison Project (S-RIP), and will be reported in Chapter 6 of the report. We had intended to highlight how the annual cycle of annular mode variability in the austral hemisphere is quite different from the boreal hemisphere, and have made an effort to better emphasize this in the revised manuscript.

Also, to better investigate the SAM, we now compare variability on monthly time scales with a station based index, established by Marshall (2003). We hope this provides a better balance of analysis of the austral hemisphere.

It seems curious that satellite observations are necessary for representing the SAM but not the NAM, one wonders why.

We speculate that this is due to the sparsity of conventional observations in the Southern Hemisphere, particularly in the higher latitudes.

P2 L20 The correlation between jet responses to global warming and annular mode persistence was not clear in the CMIP5 models (Simpson and Polvani 2016). In fact, since the annular modes are not system modes, it is possible that the overly persistent annular more timescales in comprehensive climate models may have no implication for their response to global warming (Sheshadri and Plumb 2017).

This is a good point, and we modified the introductory text here to be a bit more circumspect, and more importantly include a more comprehensive discussion of these issues in section 8.2, where both of these references are acknowledged.

We believe that the link was more useful in assessing CMIP3 and CCMVal2 (the Chemistry Climate Model Validation Project, Phase 2) because there were outlying models that more radically over-estimated the annular mode time scales.

Should one reasonably expect tropospheric jet variability that extends equatorward of your 65◦ definition (e.g. Madonna et al. 2017; Woollings and Blackburn 2012)?

Yes! As the second reviewer also had concerns about this simplified definition, we have expanded the discussion and include a new figure that shows how the polar-cap definition captures variability in the midlatitude (and tropics, at upper levels). The polar cap average does capture much of the variability of flow equatorward of 65 degrees, due to the dipole nature of the annular modes. Figure 9 of Baldwin and Dunkerton (2009) shows that correlation between the polar cap average geopotential is greater than 0.95 with the full NAM at all levels, and greater than 0.99 in the stratosphere.

Fig. 3 Has this been smoothed? If so, it would be useful to people will might try to replicate this result to mention the details.

There is no smoothing, other than taking the daily mean. It looks smoother in part because we chose a stratospheric level (10 hPa). At lower levels, you see more variability on synoptic scales.

P6 L18 Do you have theories as to why this would be the case?

This is related to the reviewers question above about why satellite measurements are needed in the austral hemisphere. We did note that JRA-55C appears to better capture the annular modes at upper levels than the surface: this may be due to the fact that annular mode widens with height, which is now explicitly shown in Figure 2. Thus, radiosondes from the tropics and midlatitudes have more value in constraining the annular mode.

Note that our new comparison with the Marshall (2003) station base index dives into this issue in more detail. The surface-input reanalyses do a better job than the full-input reanalyses. We suspect that the latter were optimized to utilize satellite measurements, and so less able to make use of surface data. The surface-input reanalyses also make use of historic data that does not necessarily get into the full-input reanalyses.

P9 L4 How do you deal with final warming (FW) events while defining strong and weak vortex events?

For consistency with previous work, we tried to follow Baldwin and Dunkerton (2001) procedure. It is possible that we have captured a few of the "dynamically" induced final warmings with our definition, which do not influence the result. (In years where the vortex decay slowly, the annular mode index does not reach sufficiently extreme values to trigger an event.)

It might be useful to examine the downward influence of FWs in both hemispheres across reanalyses, particularly since this might be of relevance to the effects of the ozone hole on tropospheric circulation in the SH.

The influence of the ozone hole on the austral circulation has been investigated as part of the SPARC Reanalysis Intercomparison Project, and will be available soon.

FWs are an aspect of stratosphere-troposphere coupling that have conventionally been studied using annular modes, that this manuscript completely ignores. Also, in general, stratosphere-troposphere coupling is thought to be strong at the end of winter and into early spring in the SH.

As noted earlier, we avoided the SFW in part because much of the literature (Black and McDaniel 2007a,b and Sheshadri et al. 2013), argued that the downward influence was not optimally captured by the annular modes. There is also less consistency on the precise

definition of the final warming. The papers above defined it by a reversal of the winds at 50 hPa and 70 degrees latitude, while in Butler and Gerber 2018, a new definition using the winds at 10 hPa and 60 N was proposed (largely in an effort to make it more consistent with the definition of Sudden Stratospheric Warmings).

An issue for this paper is that all of these definitions of the final warming depend on the zonal winds, and not the annular mode index. SFWs are being explored in the SPARC reanalysis intercomparison project, and will be reported on there. All this said, in the revision we will make a note that final warmings are important signal, especially in the austral hemisphere, where annular mode variability is concentrated at the end of the winter and spring. We decided to reference Black, McDaniel, and Robinson (2006), as this is the earliest reference on these issues (to our knowledge).

P9 L19 It is somewhat inaccurate to refer to it as a downward propagating signal, as the word propagation is typically associated with the propagation of waves. Downward "influence" or "migration" might be a better choice.

We agree that this signal is not associated with the downward propagation of a wave. To avoid confusion, we use the term migration in the revised manuscript.

P11 L2 There have been follow-up studies on the eddy feedback arguments of Lorenz and Hartmann that might be worth mentioning (e.g. Byrne et al. 2016), as these indicate that the persistence of the first annular mode might not really be the right way to think of eddy feedbacks.

We agree that these follow up studies should be discussed, and have revised the introduction to section 8.2 accordingly.

P11 L6 Could you expand on how timescales are computed, since you do not use an EOF-based definition of the annular mode?

We applied the same procedure as in Baldwin et al. 2003. The only difference is that we used our simplified annular mode index, as opposed to the EOF based index. (In a large number of calculations that I did with CMIP5 models -- which I unfortunately never published -- I found that the time scales were quite similar using both indices.) As Baldwin et al regrettably don't really explain their method, I've tried to give the gist of it in the revised text.

Response to reviewer 2

1. Page 3, line 30. There have been many different definitions of the annular mode in addition to the polar cap averaged geopotential height used in this study. Baldwin and Thompson (2009) compared different annular mode definitions. In analogy with Baldwin and Thompson (2009), it would be beneficial to briefly discuss the strengths and weaknesses of the particular annular mode used in this study.

A practical issue for this study was handling the volume of all the reanalysis data: this simple definition in part made it possible to compare all the reanalyses with an efficient and reproducible framework. Given that there are a number of definitions of the annular mode that are effectively equivalent, we also wanted to emphasize the simplest one for future researchers.

To better justify this definition, we have now added a new figure (Fig. 2) which shows that the simple polar cap average definition captures the same patterns of variability as more complicated methods.

2. Figure 3. How is the consistency determined? This does not appear to be clearly defined in either the text or the figure caption. Does the consistency correspond to the average of the six pairwise correlations between the four Reanalysis datasets?

Yes, by consistency, we meant to refer to the average pair-wise correlations between the four most modern reanalyses. We clarified the figure caption in the revision.

3. Page 6, line 19. What is meant by conventional observations? In the abstract, the term conventional appears to correspond to surface observations which doesn't appear to be consistent with the rest of the paragraph. Since on page 3, line 22, where it is indicated that JRA-55C data lacks satellite measurements, it appears that conventional in this paragraph corresponds to the exclusion of satellite measurements.

Yes, conventional observations refer to everything but satellite based measurements. In the revision, will make this distinction more clear, especially in the abstract, section 2 and in the revised Table 1. In particular, we use more clear language, distinguishing full-input, conventional-input, and surface-input reanalyses.

4. Page 7, line 6. It is stated that the NAM is consistently represented prior to the satellite era. However, this appears to be the case only for the troposphere. For the stratosphere, the ERA-20C, as indicated in the text, and the two versions of 20CR are much poorer?

It was our intent to indicate that the NAM in consistent in the full-input reanalyses; for the surface-input reanalyses, there is only consistency in the troposphere. This has been clarified in the revision.

5. Page 8, line 8. The relatively high R2 value between the 20CR and ERA-20C is mentioned for the early half of the 20th century. This is taken to indicate that the NAM may be reasonably accurate during this time period. It is not indicated in this paragraph that this result applies only for 1000 hPa.

Yes, this refers only the surface level, as stated in the text.

6. Page 9, line 11. It is stated that the strong vortex build up is less abrupt than its decay. I don't see this in Fig. 8.

We are afraid that the non-linear color bar (and potentially visual differences between the cool and warm colors) in Fig. 8 gave this incorrect impression. For weak vortex events, the index drops by over 3 standard deviations in 10-15 days, the bulk of the drop in the last 5 days. For strong vortex events, the increase in the index is only 1.5 standard deviations over approximately 40 days. We clarify this in the revision.

7. Page 9, lines 25-31. I did not follow how the sampling uncertainty is determined. Was the standard deviation determined for all four modern Reanalysis datasets over the lag days? A little more detail would be helpful.

The sampling uncertainty was determined from JRA-55 alone, and quantified by the standard deviation of the composite mean in Figure 8 a,b: it is the inter-event standard deviation (shown in Fig. 8c,d) divided by the square root of the number of events. The inter-event uncertainty for the other reanalyses is comparable to JRA-55 over the satellite era (as evidenced by the reanalysis uncertainty). As there are fewer events over the satellite period, however, there is greater uncertainty in the composite mean. This is clarified in the revised text.

8. Page 11, line 12. To remove the interannual variability, one could simply apply a high pass filter to the data with a cutoff period shorter than one year and longer than the longest e-folding time scale in the raw data.

The use of decadal means was chosen primarily to enable us to quantify the sampling uncertainty. (That is, we assume that each decade was independent, and use the differences between decades as a crude measure of the sampling uncertainty.) This is clarified in the revision.

Quantifying the variability of the annular modes: Reanalysis uncertainty vs. sampling uncertainty

Edwin P. Gerber[1] and Patrick Martineau[2]

[1]Courant Institute of Mathematical Sciences, New York University, 251 Mercer Street, New York NY 10012, USA
[2]Research Center for Advanced Science and Technology, University of Tokyo, Japan

Correspondence: Edwin P. Gerber (gerber@cims.nyu.edu)

Abstract. The annular modes characterize the dominant variability of the extratropical circulation in each hemisphere, quantifying vacillations in the position of the tropospheric jet streams and the strength of the stratospheric polar vortices. Their representation in all available reanalysis products is assessed. *Reanalysis uncertainty* associated with limitations in the ability to constrain the circulation with available observations, i.e., the inter-reanalysis spread, is contrasted with *sampling uncertainty* associated with the finite length of the reanalysis records.

It is shown that the annular modes are extremely consistent across all modern reanalyses during the satellite era (1979 onward). Consequently, uncertainty in annular mode variability, e.g., the coupling between the stratosphere and troposphere and the variation in the amplitude and time scale of jet variations throughout the annual cycle, is dominated by sampling uncertainty. Comparison of reanalyses based on conventional (i.e., non-satellite) or surface observations alone with those using all available observations indicates that there is limited ability to characterize the Southern Annular Mode (SAM) in the pre-satellite era. Notably, prior to 1979, surface-input reanalyses better capture the SAM at near surface levels than full-input reanalyses. For the Northern Annular Mode, however, there is evidence that conventional observations are sufficient, at least from 1958 onward. The addition of two additional decades of records substantially reduces sampling uncertainty in several key measures of annular mode variability, demonstrating the value of more historic reanalyses. Implications for the assessment of atmospheric models and the strength of coupling between the surface and upper atmosphere are discussed.

1 Introduction

The annular modes characterize the dominant internal variability of the extratropical atmosphere (Thompson and Wallace, 2000). In the troposphere, they primarily characterize meridional shifts in the extratropical jet stream, a positive index indicating a poleward shift. The jet stream is associated with the extratropical storm tracks, such that the annular mode indices ~~indicated~~ indicate shifts in storm activity, particularly in Northern Europe and eastern North America (e.g., Thompson and Wallace,

1998). In the stratosphere, the annular modes chiefly characterize variations in the strength of the polar vortex, a positive index indicating a stronger than average vortex. Baldwin and Dunkerton (2001) revealed a downward influence of the stratospheric polar vortex on the tropospheric jet stream by comparing annular mode indices computed at different pressure levels.

In this study, we focus on the representation of the annular mode indices, or time series, in all available global reanalysis products. The spatial structure of the annular modes is characterized by a meridional dipole, where mass (or equivalently, momentum, as the large scale flow is geostrophically balanced) is exchanged between the high and low latitudes (Thompson and Wallace, 2000). It ~~is~~ can be understood as the gravest mode of a varying jet (e.g., Vallis et al., 2004; Gerber and Vallis, 2005) and is well captured across a range of models (e.g., Gerber et al., 2010). The term "annular mode" was derived from the annular structure of the patterns in longitude, where geopotential height varies in the same phase at a given latitude (Limpasuvan and Hartmann, 2000). As discussed by Deser (2000) and Ambaum et al. (2001), among others, such coherent fluctuations are generally not observed, and the annular structure is primarily statistical in nature (Gerber and Thompson, 2017). The temporal variability of the annular modes provides a convenient means to quantify coupling between tropospheric jets and the stratospheric vortices, and variations in the amplitude and persistence of the variability through the annual cycle. It has proven more challenging to understand and model, and is thus the focus of this study.

The jet streams are known to vacillate on slower time scales than individual weather systems, as first quantified with the "zonal index" (e.g., Rossby, 1939; Namias, 1950). A number of studies have investigated how this enhanced persistence results from a feedback between eddies and the mean flow (e.g., Robinson, 1996; Feldstein and Lee, 1998; Lorenz and Hartmann, 2001; Gerber and Vallis, 2007; Barnes et al., 2010; Zurita-Gotor et al., 2014). The stratospheric polar vortex exhibits greater memory than the troposphere, with implications for predictability of the tropospheric jet stream on sub-seasonal to seasonal time scales (e.g., Baldwin et al., 2003; Sigmond et al., 2013). Models, both idealized (e.g., Gerber et al., 2008b) and comprehensive (e.g., Gerber et al., 2008a), tend to exhibit too much persistence, with possible implications for their projected response to global warming (Kidston and Gerber, 2010; Barnes and Hartmann, 2010; Son et al., 2010). It is thus of both practical and theoretical interest to determine the limits to which we can quantify the temporal variability of the annular modes of our atmosphere with reanalyses.

After briefly describing the available reanalyses and our procedure for evaluating the annular modes in Sections 2 and 3, we explore annular mode variability in the most recent products of the four major reanalysis centers in Section 4, establishing a "Reanalysis Ensemble Mean" as a standard for comparison with all reanalysis products. In sections 5 and 6, we compare the reanalyses in the satellite and pre-satellite eras, respectively. We find that the Northern Annular Mode (NAM) can be well constrained with conventional observations, and appears consistent across a number of reanalyses since 1958. Characterization of Southern Annular Mode (SAM) variability, however, appears to depend more strongly on satellite observations, and the SAM index varies substantially between ~~reanalysis records~~ reanalyses before 1979. There is evidence, however, that surface-input reanalyses can capture the tropospheric SAM, at least on monthly time scales, as far back as the late 1950s. In Section 7, we further explore historic reanalyses constrained only by surface measurements, and find that there is potential to constrain the ~~annular modes in the Northern Hemisphere for~~ NAM at tropospheric levels in earlier periods.

~~Given the state of the reanalyses, in~~ In Section 8 we show that sampling uncertainty is the leading source of uncertainty in our ability to constrain the temporal variability of the annular modes. We focus in particular on the downward coupling associated with extreme vortex variability and the time scales of annular mode variability as a function of height and season. Our conclusions are summarized in Section 9.

2 The reanalyses

Table 1 lists the reanalyses compared in this study. Fujiwara et al. (2017) provide a detailed explanation of the data assimilation procedures and input data sets that differentiate these reanalysis products~~, but~~. Here, we highlight a few key differences ~~here~~that are pertinent to this study. ERA-Interim, JRA-55, MERRA2, and ~~NCEP-CFSR~~ CFSR/CFSv2 represent the current state-of-the-art reanalysis products of the four major reanalysis centers. While this is an moving target, e.g., the new ERA5 reanalysis will soon replace ERA-I as the top product available from the European Center for Medium Range Weather Forecasting (ECMWF), a key finding of this ~~paper~~ study is that all of the modern reanalyses accurately capture the variability of the annular modes.

In constructing a *re*analysis, one seeks a balance between the aim of providing a best possible estimate of the atmospheric state at any given time (as with an *analysis* for providing initial conditions for a weather forecast), with the goal of providing a homogeneous record of the atmospheric evolution. The latter requirement generally entails a choice of a fixed atmospheric model and assimilation procedure for the entire reanalysis. This is an issue with the CFSR and CFSv2 reanalyses. The National Center for Environmental Prediction (NCEP) made substantial upgrades to their reanalysis system as of Jan. 1 2011, including a change in the resolution of the atmospheric model as documented in Saha et al. (2014). We find that the annular modes appear to be continuous and consistent across this break, but it is best practice to consider CFSR and CFSv2 as separate reanalyses.

A more difficult decision concerns the use of new observations that become available during the record~~, the~~. A most notable change ~~being~~ is the introduction of satellite measurements c. 1979. ~~In general, the most state-of-art reanalysis~~ Most reanalyses make use of all observations as they become available~~, providing~~. This provides an increasingly accurate representation of the atmospheric state with time~~. This introduces a risk of~~, but introduces the potential for spurious trends which are not associated with physical changes in the atmosphere, but rather an enhanced ability to observe ~~the atmosphere.~~ it. A good example of these artificial trends can be seen in stratospheric temperatures, as shown in Fig. 1 of Long et al. (2017). In earlier reanalyses, global mean temperatures jump abruptly in response to new satellite measurements. More recent reanalyses have adopted bias correction techniques to adjust for the errors in earlier measurements, but there is still potential for spurious trends over time. There are no longer discontinuities in the temperature record, but there is still potential for spurious trends, especially in earlier years where the bias correction becomes more distant.

A few reanalyses explicitly limit the observational inputs, giving us an opportunity to assess the impact of new observations. In particular, ERA-20C and NOAA-20CR v2 and v2c intentionally limit themselves to surface based observations only~~. This~~; we refer to these as "surface-input reanalyses". This choice allows them to provide ~~an~~ a more uniform estimate of the atmospheric state as far back as 1850. They provide us a unique opportunity to assess the coupling between the surface and

interior circulation of the atmosphere, particularly the stratosphere. They are not immune to spurious trends, however, as the availability of surface measurements fluctuations with time.

The JRA-55C reanalysis uses the same model and data assimilation procedure as in JRA-55, but explicitly excludes all satellite measurements. ~~It therefore~~ We refer to it as a "conventional-input" reanalysis, but note that all reanalyses are, by this definition, conventional-input before 1979. JRA-55C provides an opportunity to assess the impact of satellite measurements on our ability to quantify the annular modes. ~~The~~ Finally, we consider JRA-55AMIP ~~"reanalysis"~~ integration, which is not a reanalysis. It only considered observed sea surface temperatures as an observational input, i.e., it is the output of a free running atmospheric model used in the JRA-55 and 55C reanalyses, constrained only by the sea surface temperature and observed changes in radiative forcing. As will be shown, it allows us to confirm the conventional wisdom that in general, the extratropical circulation is only weakly constrained by sea surface temperature.

3 The annular modes

As suggested by Baldwin and Thompson (2009), we use a simplified procedure for computing the annular mode based on the polar cap averaged geopotential height, where the cap is defined as all latitudes poleward of 65°. Following Gerber et al. (2010), we first remove the global mean geopotential height at each time step. This focusses the analysis on meridional variations in geopotential height (i.e., shifts in mass), as opposed to systematic variations in the entire layer associated with changes in atmospheric temperature. In practice this only has a significant impact on the indices in the upper stratosphere and above, but it also helps removes trends associated with global warming. Based on model integrations (e.g., Gerber et al., 2010), ~~trends can become~~ we expect that trends will be most significant near the tropopause, where there is a sharp gradient in the character of the annular mode as it shifts from characterizing the tropospheric jet to the stratospheric vortex.

Fig. 1 illustrates our procedure, which we spell out precisely in the following.

1. 6 hourly output of geopotential height is averaged to form a daily time series $Z(t,\lambda,\phi,p)$, where λ, ϕ, and p refer to latitude, longitude, and pressure. For leap years, the 366 days are interpolated to 365, but we note the impact of this interpolation is inconsequential, and one could more simply remove one day from each leap year.

2. The annual cycle is formed by averaging each calendar day over the record, and then smoothing the resulting annual cycle with a 60 day low pass Lanczos digital filter. The anomalous height $Z'(t,\lambda,\phi,p)$ is formed by removing the smoothed annual cycle. The daily, zonal mean anomalous height is illustrated for a single year and pressure level in Fig. 1a.

3. We then compute the global mean geopotential height, $\overline{Z'}^{global}$, and austral and boreal polar cap heights, $\overline{Z'}^{SH}$ and $\overline{Z'}^{NH}$, as illustrated in Fig. 1b.

4. Raw SAM and NAM time series are defined by $-(\overline{Z'}^{SH} - \overline{Z'}^{global})$ and $-(\overline{Z'}^{NH} - \overline{Z'}^{global})$, respectively. The negative sign is in keeping with the standard convention of the annular mode introduced by Thompson and Wallace (1998).

5. The standardized SAM and NAM indices are obtained by dividing the raw indices by their standard deviations, yielding indices with unit variance by construction, as illustrated in Fig. 1c.

As steps 1-4 are linear, one can rearrange them to minimize the computational effort. In practice, we first compute the global and polar cap average geopotential heights, and then daily average and deseasonalize them to form time series $\overline{Z'}^{global}$, $\overline{Z'}^{SH}$, and $\overline{Z'}^{NH}$.

~~The~~ While this simple definition of the annular mode index appears to depend only on geopotential height over the pole, we stress that it captures variability throughout the midlatitudes. Fig. 2 illustrates the latitudinal structure of the annular mode variability, obtained by regressing daily geopotential height anomalies on a given level on the respective annular mode index, i.e., the 850 hPa patterns are obtained using the 850 NAM and SAM indices, and so forth. It compares remarkably well with the structure obtained in EOF calculations (e.g. Gerber et al., 2010, their Figure 4). The broad latitudinal extent of the patterns reflects the strong anticorrelation between geopotential height in the high and low latitudes, as necessitated by the conservation of mass (cf. Gerber and Vallis, 2005). One can even obtain the zonal structure of the annular mode patterns by regressing two dimensional (latitude-longitude) geopotential on the annular mode index (not shown): in the Northern Hemisphere, the zonal mean geopotential anomalies are dominated by localized variability in the Atlantic and Pacific storm tracks.

By convention, the AM indices are normalized to have unit variance. The patterns thus capture the amplitude of variations. In the troposphere, variability is stronger in the austral hemisphere, as seen in the increased amplitude of the SAM relative to the NAM at 850 hPa. In the stratosphere, however, variability is stronger in the boreal hemisphere, as seen with the 100 and 10 hPa patterns.

We took the opportunity in Fig. 2 to contrast the patterns obtained from a full-input reanalysis (ERA-Interim, which is characteristic of the other modern reanalyses) with those in a surface-input reanalysis (ERA-20C). The agreement is weaker at higher altitudes, consistent with a decorrelation between the indices at height (see Section 5), but we find it notable that a reanalysis which only assimilates surface information can reasonably capture the structure of annular mode variability at 10 hPa. This is in part a reflection of the fact that the dipolar structure of the annular modes is fairly generic, and well obtained by any free running atmospheric model.

Note that the formation of the annual cycle and normalization of the time series (steps 2 and 5) depend on the length of the record. If one follows this procedure to define the annular mode over two distinct periods that overlap, the resulting indices will not agree perfectly during the period of overlap. We find that that difference is inconsequential provided one uses a period of sufficient length, a decade or so in practice. This limitation could be addressed by defining the climatology and normalization constants over a set period, for instance the World Meteorological Organization (WMO) climatology spanning the last three full decades. For the purposes of this paper, we have computed the indices separately for each period of comparison, e.g., ~~1980-2016~~ 1981-2010 for our comparison of the ~~modern reanalyses in Section 4 , or for 1981-2010 for~~ reanalyses in Sections 4 and 5, and 1958-1978 for our comparison of ~~all reanalyses in Section 5~~reanalyses in the pre-satellite era in Section 6.

We focus on a subset of the pressure levels that were shared by all reanalyses: 1000, 850, 700, 600, 500, 400, 300, 250, 200, 150, 100, 70, 50, 30, 20, 10, 7, 5, 3, 2, and 1 hPa. Levels above 10 hPa are unavailable for NCEP-R1/R2 and NOAA 20CR

v2/v2c reanalyses. The annular mode indices are fairly uniform within the troposphere and stratosphere, respectively, and such fine vertical resolution is really only needed in the tropopause region.

Lastly, we ~~noted~~ remark that our definition of the annular mode (or equivalently, previous Empirical Orthogonal Function based definitions) require extrapolation of data below the surface. This was done by the reanalysis centers for all reanalyses with the exception of MERRA products. We have opted to omit MERRA and MERRA2 from comparisons below 700 hPa, where their data was incomplete.

4 Consistency in the representation of the annular modes in state-of-the-art reanalyses

We first compare the four modern reanalyses, ERA-Interim, JRA-55, MERRA2, and NCEP-CFSR, to justify the use of "Reanalysis Ensemble Mean" (REM) annular mode indices as a benchmark of comparison. ~~For a comprehensive comparison, we used the full period shared by all of these reanalyses, 1980-2016. Note that the NCEP-CFSR reanalysis was upgraded substantially in 2010~~We focus on the period 1981-2010, which corresponds to the standard 30 year climatological evaluation period by WMO convention. This also conveniently avoids the break between the CFSR and CFSv2 reanalyses, but we ~~see no evidence of a discontinuity in the annular mode variability across this break~~assess this transition below, in addition to preliminary output from the new ERA5 reanalysis.

Fig. 3 shows the pairwise squared correlation (R^2) of the NAM and SAM indices as a function of pressure. The squared correlation coefficient indicates the fraction of variance shared by ~~the~~ two time series. We find that the modern reanalyses share approximately 96% of the variance or greater at all levels in both hemispheres, with the exception of the upper stratosphere in the austral hemisphere.

In the Northern Hemisphere, the agreement is 99% or better between ERA-Interim, JRA-55, and MERRA2, with slightly weaker correlation to ~~NOAA-CFSR~~ CFSR in the upper troposphere and lower stratosphere, where there is a minimum in correlation at 200 hPa near the extratropical tropopause. ~~NOAA-CFSR~~CFSR, however, does not stand out from the other reanalyses at this level in the austral hemisphere. There is a tendency towards better agreement in the lower and ~~mid stratosphere~~ mid-stratosphere as compared to the troposphere in both hemispheres, but we stress that the correlation is always extremely high.

Given the strong agreement of all the modern reanalyses, we adopted a REM annular mode index based on the average of ERA-Interim, JRA-55, and ~~NOAA-CFSR~~CFSR. MERRA2 was omitted from the REM due to missing data at lower levels, but the results are nearly identical if it is included. We emphasize that this decision does not imply a value judgement of the quality of MERRA2 reanalysis.

The intensity of annular mode variability changes throughout the annual cycle, particularly in the stratosphere (e.g., Baldwin et al., 2003; Gerber et al., 2010). This can be seen in the annual cycle of annular mode variance in Fig. 4 (black contours). In the troposphere, the annular modes are most variable in the winter seasons in both hemispheres, with a greater variation over the annual cycle in the boreal hemisphere. In the stratosphere, there is considerably more variation throughout the annual cycle: the maximum is approximately 4 times the annual mean, and variability drops to near zero in the summer. There is also

a notable difference in the timing of peak variability between the two hemispheres. For the NAM, the maximum is collocated with the ~~troposphere in~~ maximum in tropospheric variability in the winter, while for the SAM, variability is greatest in late spring.

Given this variation of the annular modes within the annual cycle, we check the consistency between the reanalyses as a function of day of year, illustrated by the shading in Fig. 4. We find that the indices agree well ($R^2 > 0.95$) at all levels and heights except when the variability is weak during the summer, where on average, the pairwise R^2 drops as low as 0.5. Thus, there is great certainty in the annular mode state except in periods when the annular mode is inconsequential.

Finally, we consider the consistency of the annular mode indices in more recent years. NCEP made substantial upgrades to their reanalysis system starting in 2011, replacing the CFSR reanalysis with CFSv2. To assess the new system, in Fig. 5 we compare the R^2 correlation between the SAM in CFSR with other reanalyses between 2005-2010 with the same metric for CFSv2 for the period 2011-2016. This choice of periods provided the most fair comparison: 6 years for each reanalysis. We find that there is a reduction in the spread between CFSv2 and the other reanalyses relative to CFSR, most notably at upper levels, where the R^2 correlation increases to approximately 99% from 95%. As seen in Fig. 5b, however, coherence is increased at all levels – although it is important to point out that CFSR was already highly correlated with the other reanalyses. We have not shown the equivalent plots for the NAM because we found no discernable difference between the two periods; both plots were equivalent to Fig. 3a. This is in part because the NAM in CFSR was already highly correlated at all levels and their was more room for improvement with the SAM. We have additionally inspected the annular indices across the break from CFSR and CFSV2 at 31 December 2010, and found no abrupt changes (not shown).

ECMWF is in the process of producing a new reanalysis, ERA5. We have found ERA5 to be virtually indistinguishable from the other modern reanalyses over the period for which it is currently available, 2008-2016. During this interval, the NAM and SAM based on ERA5 output are R^2 correlated at 0.998 or better with those in ERA-Interim and MERRA2 at all available levels. For JRA-55, R^2 exceeds 0.996 (0.991) for the NAM (SAM); the SAM coherences is only less than 0.996 near the surface, where interpolation is likely the source of error. Care must be taken in comparing with CFSR/CFSv2, given the transition at the end of 2010, and Fig. 5 shows the high correlation with ERA5 over the period 2011-2016. ERA5 should ultimately be available from 1950 onward, and it will be interesting to explore the pre-satellite era in this reanalysis.

5 Comparing the representation of the annular mode indices across all reanalyses

Fig. 6 shows the R^2 correlation between the annular mode indices computed from each individual reanalysis with the REM annular mode index over the period 1981-2010. This period ~~, which corresponds to the standard 30 year climatological evaluation period by WMO convention, also provides~~ provides nearly optimal overlap between the reanalyses: all but one (ERA-40) are available the entire period. The correlation for ERA-40 was based on the years when it was fully available, 1981-2001. We enumerate the key findings below.

(1) The NAM is well captured throughout the troposphere and the bulk of the stratosphere by all reanalyses that assimilate free atmospheric data. The R^2 correlation with the REM exceeds 0.975 from the surface to 10 hPa for all reanalyses excepting

ERA-20C, JRA55AMIP, and NOAA 20CR v2/2c. The modern reanalyses are somewhat better correlated with the REM than earlier generation products, particularly in the lower to mid-stratosphere. While this could partly arise by construction (the REM is based on the modern reanalyses), MERRA and MERRA2 are not a part of the REM and still exhibit enhanced correlation. MERRA appears slightly more highly correlated than MERRA2, although the difference is trivial.

(2) The SAM is well captured by most reanalyses that assimilate satellite measurements, and is demonstratively better represented in the ~~"second generation"~~ more recent reanalyses. Most of the comprehensive reanalyses are well correlated with the REM ($R^2 > 0.92$ up to 10 hPa), but in comparison to the Northern Hemisphere, the modern reanalyses are consistently better than earlier reanalyses; MERRA and MERRA2 are better correlated with the REM that the ~~first generation~~ earlier reanalyses (ERA-40, JRA-25, ~~NOAA-R1~~ NCEP-R1 and R2) at all levels.

(3) Satellite measurements are critical for the representation of the SAM, but not for the NAM. The representation of the NAM by JRA-55C, which assimilates only conventional observations, is nearly indistinguishable from the second generation reanalyses up to 30 hPa, and is still very good up to 10 hPa. We conclude that conventional observations are sufficient until the upper stratosphere, which suggests that there is potential for a skillful representation of the NAM in the pre-satellite period, as further investigated in Section 6. In the austral hemisphere, however, the representation of SAM in JRA-55C is demonstrably poorer at all levels (though it still captures more than 85% of the variance up to 3 hPa). JRA-55C is most strongly correlated with the other reanalyses in the mid-stratosphere, but decays considerably towards the surface, where it only captures 85% of the variance, substantially less than the reanalyses which assimilate only surface measurements.

(4) Reanalyses based on surface measurements alone remain highly correlated with the REM in the troposphere, but the correlation falls off sharply above the tropopause, particularly for the 20CR reanalyses. ERA-20C, however, remains remarkably well correlated through the stratosphere, capturing more than half the variance of the REM up to the stratopause at 1 hPa. As discussed in greater detail in Section 7, this indicates that the stratosphere is sufficiently coupled to the surface to extract useful information about its state from surface data alone.

We finally note that sea surface temperatures are insufficient to capture AM variability, at least in the case of the JRA-55AMIP integration. It has long been known that the extratropical circulation is not driven by local SSTs (e.g., Barsugli and Battisti, 1998). Tropical SSTs are known to influence the annular modes (e.g., El Niño; L'Heureux and Thompson, 2006), but appear insufficient to constrain the annular mode variability of either hemisphere.

6 Pre-satellite representation of the annular modes

The ability of the JRA-55C to accurately capture the NAM without the aid of satellite measurements suggests that it should be possible to ~~assess~~ capture it during the pre-satellite era~~.~~, at least with full-input reanalyses that make use of all available observations. (We note that prior to 1979, there's no distinction between a full- and conventional-input reanalysis.) Fig. 7 compares the ~~5~~ 6 available reanalyses during the period 1958-1978. As it is less trivial to identify a meaningful REM during this period, we consider two comparisons against a given individual reanalysis, JRA-55 and NCEP-R1, respectively. Comparable results are found when using ERA-40 as a reference, and the selection here is not meant to be a value judgment. Based on

Fig. 7a and b, we concluded that the NAM is consistently represented throughout the troposphere and stratosphere in ERA-40, JRA-55, and NCEP-R1, giving us confidence that there are sufficient observations to quantify the NAM as far back as 1958.

While the NAM in the ~~surface-only~~ surface-input reanalyses remains highly correlated in the troposphere, we note that ERA-20C is not as well correlated in the stratosphere as it was in more recent years. At 10 hPA, ERA-20C only captures about 40% of the variance, as compared to 60% between 1981 to 2010.

The situation is quite different in the austral hemisphere, with widespread divergence between the reanalyses (Fig. 7c,d). JRA-55 shares only 30% of the variance with the other reanalyses throughout the troposphere. NCEP-R1 is more strongly correlated with the NOAA 20CR and ERA-20C reanalyses in the troposphere (sharing approximately 60% of the variance), but poorly correlated with the other full reanalyses (JRA-55 and ERA-40). We conclude that the SAM is poorly constrained in the pre-satellite era. This is partly expected, given the limited ability of JRA-55C to capture the SAM during the satellite era. JRA-55C was still fairly well correlated with the REM in the satellite era, however, suggesting that the poor representation of the SAM before 1979 also reflects a dearth of conventional observations.

The poor correlations in Fig. 7(c,d) begs the question: do any of the reanalyses have skill before 1979? It is difficult to characterize the synoptic evolution of the SAM from direct measurements, but Marshall (2003) developed a station-based index to track the SAM at the surface on monthly time scales. Briefly, the index uses a series of stations to estimate the difference in zonal mean pressure between 65 and 40°S. Table 2 compares correlation between the Marshall (2003) index over the pre- and post-satellite periods with the 850 SAM index. From 1979 and 2001, the station-based index is approximately 0.85 correlated with the near-surface SAM in the 6 reanalyses considered; similar correlations are observed with the other reanalyses (not shown). We believe that this level of correlation reflects differences in the definition of the indices, as opposed to errors in the reanalyses.

In the full-input reanalyses, correlation between the near surface SAM and the station-based index drops markedly in the pre-satellite period. In ERA-40 and JRA-55, the R value of about 0.5 indicates that that indices share only 1/4 of the variance. The correlation, however, is effectively the same over both periods for the surface-input reanalyses. This suggest that there is greater skill in these reanalyses (and in NCEP-R1, albeit less so) that the more modern full-input reanalyses. The poor behavior of the full-input reanalyses is puzzling, but may reflect the fact that these systems have been optimized for the satellite era, and so less capable of working with more limited observations. (The surface-input reanalyses may also make use of additional historical observations that are not incorporated in the full-input reanalyses.) The consistent correlation between the SAM in the surface-input reanalyses with the station-based index suggests that we do have a good handle on near-surface variability of the SAM as far back as 1958, at least on monthly time scales. This skill likely extends through the troposphere, given the barotropic nature of annular mode variability, but is less likely to extend into the stratosphere.

As seen in Fig. 6d, the SAM in JRA-55C is better correlated with the REM in the stratosphere than it is in the troposphere during the satellite era. Similarly puzzling behavior is observed in the pre-satellite era: reanalyses which do take in measurements from the free atmosphere are better correlated in the stratosphere than below. We speculate that this is due to the broadening of the spatial structure of the SAM with height~~(e.g., Gerber et al., 2010, their Fig. 4).~~. As seen in Fig. 2, the pattern of annular mode variability at 10 hPa extends all the way into the tropics. Similar broadening is observed at other stratospheric

levels (not shown), and is already discernable at 100 hPa level in the austral hemisphere (Fig. 2a). Consequently, observations from the midlatitudes and tropics are sufficient to constrain the annular mode at higher elevations, and reanalyses do not suffer from limited observations over Antarctica.

7 Representation of the annular modes in ~~surface-only~~ surface-input reanalyses

We explore the ability of reanalyses that only incorporate surface observations to capture annular mode variability in Figs. 8 and 9. The former compares two years of the NAM indices represented by ERA-20C with those of the REM at 1000, 100, and 10 hPa. As indicated in Fig. 6a, ERA-20C captures approximately 80% of the variability of the NAM at 100 hPa, and 60% of the variability at 10 hPa during the satellite era. The accuracy of ERA-20C appears to fluctuate from year to year. In this example, the winter of 2007-8 is captured with remarkable fidelity, while during the winter of 2006-7 there is an overall negative bias at upper levels, particularly at 10 hPa, although much of the high frequency variability is still captured.

This specific period was selected in part to highlight the ability of the ERA-20C to capture Stratospheric Sudden Warming events, marked by the yellow circles in the 10 hPa indices. ERA-20C captures the two events over this period, at least within a few days of the modern reanalyses. The annular modes in the NOAA 20CR reanalyses (not shown) are comparable to ERA-20C in the troposphere (ERA-20C doing slightly better in the Northern Hemisphere, NOAA-20CR slightly better in the Southern Hemisphere), but NOAA-20CR struggles to capture variability above the tropopause (Fig. 6). As found by Butler et al. (2017), the polar vortex in NOAA-20CR v2c is too strong (or equivalently, exhibits a cold bias), and exhibits only one major warming event between 1958 and 2011.

In Fig. 9, we compare the evolution of correlation between the annular mode indices in ERA-20C with other reanalyses that extend into the pre-satellite era. An 11 year moving window allows us to observe how the correlation changes over time. Focussing first on the boreal hemisphere, we find that the NAM at the surface (solid lines) is well constrained in all the reanalyses from 1950 on. The R^2 correlation with NOAA-20CR is relatively high throughout the record, albeit decaying to about 60% by the start of the century. This suggests that there may be sufficient observations in the first half of the century to capture the majority of annular variability in the troposphere.

The correlation between the ERA-20C NAM at 100 and 10 hPa with conventional reanalyses NCEP-R1 and JRA-55 ~~does appear~~ appears to weaken as one moves back in time. At 10 hPa, the correlation drops noticeably around 1975. We have no means to assess the skill of NCEP-R1 before 1958. Hence the drop in correlation with ERA-20C around 1950 could be due to a loss in skill in either (or both) reanalyses. As noted ~~above~~earlier, the NOAA-20CR reanalyses exhibit limited skill in the stratosphere, and correlation with ERA-20C is low at all times.

In the Southern Hemisphere, the 1000 hPa annular mode indices in ERA-20C and the NOAA-20CR reanalyses are well correlated ~~to about~~ with each other as far back in time as 1950, ~~at which point correlation falls off to~~ before which point the correlation is near zero. ~~We conclude there is insufficient information to constrain the austral circulation before this date.~~

~~ERA-20C diverges from the conventional reanalysis~~ As found from the comparison with a station-based index in Table 2, we have reason to trust tropospheric SAM indices in the surface-input reanalyses from at least the late 1950s onward. Before

1950, we likely have insufficient information to make an informed estimate. The full-input reanalyses (JRA-55 and NCEP-R1) ~~earlier, but recall that JRA-55C (which is equivalent to JRA-55~~ are not well correlated with ERA-20C before 1980, but given that they diverge from the station-based SAM in the pre-satellite era~~) had difficulty capturing the annular mode during the satellite era. We thus have little~~ , we have reason to trust ~~the conventional reanalysis over~~ ERA-20C ~~and NOAA-20CR. ERA-20C appears to capture much of the variability at higher levels until about 1980, at which time the correlation drops off abruptly. As discussed in the previous section, we~~ over them. We have little reason to trust estimates of the SAM above the tropopause in any of the ~~conventional reanalyses above the surface in the~~ reanalyses in the pre-satellite era. It is interesting, however, that their appears to be non-trivial (albeit weak) correlation between the SAM in ERA-20C and those in NCEP-R1 and JRA-55 at upper levels in the 1960s, more so than observed during the 1970s.

8 Sampling uncertainty vs. reanalysis uncertainty

We conclude our study by focusing on the factors limiting our ability to quantify the behavior of the annular modes in our atmosphere. We contrast uncertainty associated with the finite length of the observational record, or "sampling uncertainty" with spread in the statistics between the reanalyses, or "reanalysis uncertainty". We focus exclusively on the most modern reanalyses, but the results from Section 5 suggest that nearly all the reanalyses capture variability of the annular modes fairly well during the satellite era.

8.1 Coupling between the stratospheric polar vortices and tropospheric jets

We first consider the uncertainty in the downward coupling between the stratosphere and troposphere associated with weak and strong vortex events, as explored by Baldwin and Dunkerton (2001). Fig. 10a,b illustrates composites of the annular mode indices as a function of height for weak vortex events in the JRA-55 reanalysis, chosen because it offers the longest record (1958-2016). Weak events, associated with Sudden Stratospheric Warmings (SSWs), and strong events are defined as in Baldwin and Dunkerton (2001): when the annular mode index at 10 hPa dips below (rises above) -3.0 (1.5) standard deviations, respectively. ~~A~~ Additionally, a 30 day seperation is required between events. We note that these criteria largely leave out Stratospheric Final Warmings. As found by Black et al. (2006), the final break down of the vortex is often event-like, but does not project as well onto the annular mode indices in comparison with SSWs.

The extended record affords approximately twice as many events available to Baldwin and Dunkerton (2001). The key structure of the composite remains about the same, with two notable exceptions. Following weak vortex events, the response of the tropospheric annular mode appears more abruptly, trailing ~~just by~~ by just a few days, although it still clearly lags weakening of the polar vortex in the stratosphere. The faster response at the surface is consistent with composites of hundreds of events from Coupled Model Intercomparison Project, Phase 5 (CMIP5) integrations (Charlton-Perez et al., 2013, their Fig. 5).

The coupling between the troposphere and stratosphere also appears to be more simultaneous for strong vortex events, which are themselves notably less "event-like". This fact is partially obscured by the nonlinear color scale in Fig. 10a,b. In a weak vortex event, the annular mode index at 10 hPa drops 3 standard deviations – corresponding to a 1.4 km rise of the 10 hPa

surface at the pole – in less than 2 weeks, the bulk of the change occurring in the final days before the event. The build up of the polar vortex is less abrupt than its destruction in SSWs ~~,~~ – the 10 hPa index rises 1.5 standard deviations over 40 days – making the detection of strong vortex events more sensitive to the selection criteria. For example, the separation requirement plays a more important role in the definition of strong vortex events than for weak vortex events. The ~~higher~~ stronger than average vortex preceding event onset could potentially be ~~somewhat reduced~~reduced, albeit only a small amount, by requiring a longer window between events.

Fig. 10c,d shows the event-to-event standard deviation of the annular mode index as a function of height and lag. White delineates the region where the variability is near unity, equal to the climatological value. Inter-event variance is reduced at 10 hPa at the time of onset by construction, but is otherwise generally at or above the climatological value. The mean impact of events on the troposphere (i.e., the signal) is approximately 0.3 to 0.6 standard deviations in amplitude, and thus easily overwhelmed by this natural variability. Hence many SSWs will not appear to be associated with a downward ~~propagating~~ migrating signal.

Note that we have normalized the variance to be of order unity in the annual mean. As seen in Fig. 4, the variance is above average in the winter, when strong and weak vortex events occur. The average tropospheric value ~~(~~ over lags -90 to 90 is 1.2 ~~)~~ ~~is~~ standard deviations, consistent with the mean winter variability, and the enhanced variance of the polar vortex approximately 20-40 days after strong vortex events is comparable with the mean variability of the winter vortex.

In Fig. 11 we contrast the sampling uncertainty in these events (as determined from the JRA-55 reanalysis) with two measures of uncertainty associated with differences between the reanalyses. The sampling uncertainty (panels a,b) is expressed as a one standard deviation error bound ~~.~~ on the composites shown in Fig. 10a,b. It is simply the event-to-event standard deviation shown in Fig. 10c,d divided by the square root of the number of events. For both weak and strong vortex events, the sampling error in the troposphere is fairly uniform in time. For weak vortex events, the average standard error over the 181 day period surrounding events is 0.21 at 300 hPa; this level is characteristic of other tropospheric levels. The average uncertainty is 0.12 for strong events, smaller because the sample size is nearly three times as large, a consequence of the weaker event threshold. We note that the event-to-event variability in the other modern reanalyses is comparable to than in JRA-55. As they provide fewer events, however, the uncertainty of the mean composite is greater, as discussed below.

For computing the "reanalysis uncertainty" we restrict the analysis to the common period 1980-2016. Fig. 11c,d shows what we have termed the "raw" uncertainty: the standard deviation between composites of weak/strong vortex constructed independently from the four modern reanalyses. At 300 hPa, the average ~~uncertainty~~ spread between the reanalyses over the -90 to 90 day period is just 0.025 for weak events (8 times less than the ~~sample~~ sampling uncertainty) and 0.057 for strong vortex events (less than half the respective sampling uncertainty).

Much of this inter-reanalysis spread, however, arises from the fact that the composites are not necessarily built about the same events. As emphasized by ~~?, their Fig. 4~~Martineau et al. (2018a, their Fig. 4), differences between reanalyses are much smaller than differences between events. The threshold criteria is sensitive to subtle changes in the annular mode index, such that not all reanalyses indicate the same event dates. With respect to weak vortex events, 19 were identified in all the reanalyses, 13 of which fell on the same day across all of them. For one event, the date at which the 10 hPa NAM index crossed the -3 standard

deviation threshold was offset by 1 day in ERA-I reanalysis, and in 5 other events, the date in ~~NOAA-CFSR~~ CFSR/CFSv2 was offset by one day.

Given the large variability of the NAM on daily time scales, these subtle offsets matter. When the event dates are fixed across all the reanalyses, the inter-reanalysis spread is further reduced, as shown in Fig. 11e,f. Here, we again plot the standard deviation between composites of weak/strong vortex events across the reanalyses, but now using identical events dates. (The dates were set to day where the event fell in the majority of the reanalysis). For weak vortex events, the average tropospheric uncertainty is reduced to 0.016, over 50% less than in the raw composite shown in Fig. 11c, and 13 times less than the sampling uncertainty.

Strong vortex "events" are less event-like by nature, leading to greater spread in their timing between the reanalyses. Analyzing each reanalysis separately, we identified 49 events in ERA-40, 52 in JRA-55 and MERRA2, and 51 in ~~NCEP-CFSR~~CFSR/CFSv2. The spread between these composites, Fig. 11d, thus includes some sampling uncertainty (i.e. different events), and variations due to differences in timing (which were up to a week or so in many cases).

To remove this sampling uncertainty, we focus on 53 unique events which were identified by at least 2 reanalyses. There was universal agreement on the date for 20 of these events, and three reanalyses identified the same date in 11 more cases. For the other 24 cases, at least two reanalyses differed on the timing and/or missed the event all together. When composites of weak vortex events were formed for each reanalysis using the same dates (Fig. 11f), the standard deviation between reanalyses drops substantially. At 300 hPa, the average uncertainty was just 0.011, about half the level of the smallest contour on the plot, and about 10 times less than the sampling uncertainty.

For both weak and strong vortex events, then, the uncertainty associated with differences between the reanalyses is a factor of 10 less than uncertainty associated with sampling in the extended 59 year JRA-55 record. As the sampling uncertainty decays with the square of the number of events, we would need a record that is approximately 100 times longer (i.e., 6 millenia of observations) before sampling uncertainty is reduced to the same level as the reanalysis uncertainty!

As sampling uncertainty is by far the limiting factor, the fact that high quality reanalysis of the Northern Hemisphere appears possible back to 1958 is important. The addition of 21 years lengthens the record by approximately 50% percent, thereby reducing the sampling uncertainty in the JRA-55 composite by approximately 20% percent. This is to say, if we determine the sampling uncertainty from the other reanalyses (ERA-Interim, MERRA2, or CFSR/CFSv2), the sampling uncertainty is about 20% greater, due to the reduction in the number of events. It is planned to extend ERA5 back to 1950, and we look forward to comparing this reanalysis against JRA-55 in the pre-satellite era.

8.2 The time scales of the annular modes

As ~~noted in the introduction, the temporal variability of the annular modes indicates the strength of~~ far back as Rossby (1939) and Namias (1950), it has been recognized that there is greater persistence of the zonal mean circulation in the midlatitudes, as compared to synoptic conditions at any give location. Robinson (1996) and Lorenz and Hartmann (2001) suggest that this enhanced persistence could reflect the possibility of eddy-mean flow feedbacks~~in the troposphere, and~~ . While this is clearly the case in idealized models (e.g., Garfinkel et al., 2013), more recent work by Byrne et al. (2016) questions whether such

feedbacks are observed in the real atmosphere. In addition, Sheshadri and Plumb (2017) question whether the annular mode should be viewed as a stationary pattern, characterized by a single decay time scale.

A heightened sensitivity to external forcing is associated with excessive persistence of the annular modes in idealized atmosphere models (Gerber et al., 2008b) and earlier generation climate models (Kidston and Gerber, 2010; Barnes and Hartmann, 2010; S but this effect appears to be dominated by outlying models, and so less relevant to modern GCMs. Simpson and Polvani (2016) find that there is a weak correlation between the time scale of variability and jet response, but the seasonal timing is off: models with longer time scales in austral summer tend to respond more strongly in austral winter.

Capturing the correct time scales of variability has proven to be a challenging test for atmospheric models ~~.~~ (e.g., Gerber et al., 2008a). We thus ask how well it can be constrained by the reanalyses. Baldwin et al. (2003) quantify the persistence variations in jets by the e-folding time scale, τ, of the annular mode index. It can be computed as a function of height, to contrast variability in the stratosphere and troposphere, and calendar date, to capture seasonality of the jet variations. We employ the exact same procedure as in Baldwin et al. (2003), except that we apply it to our simpler, polar-cap average definition of the annular mode. Briefly, τ is computed as the best fitting e-folding time scale to the annular mode autocorrelation function from lag 0 to 50, separately at each height. It is computed separately for each day of the year, localized within the annual cycle by first applying a Gaussian filter of width 60 days at half maximum to the time series, centered about the calendar date for all years in the record.

The annular modes in the troposphere exhibit the longest time scales (Fig. 12a,b) during the seasons when the stratosphere is most variable (Fig. 4): January-February in the boreal hemisphere and November-December in the austral hemisphere. This suggests a downward influence of the stratosphere during the seasons when coupling is most active. Gerber et al. (2010) found that models share this behavior, although both the peak in stratospheric variability and in tropospheric time scale tend to reach a maximum later in the annual cycle, particularly in the austral hemisphere.

As noted by Keeley et al. (2009), the e-folding measure τ is influenced by variability on both synoptic and interannual time scales, and so unable to comprehensively characterize the temporal variability (cf. Osprey and Ambaum, 2011). To limit the influence of low frequency variations, here we compute the time scale on non-overlapping decadal periods, averaging them to produce Fig. 12a,b. A calculation based on the full record (thus allowing the influence of time scales greater than a decade) exhibits the same structure, but with greater τ values in regions more sensitive to low frequency variations and trends: the boreal stratosphere summer (when variability is weak) and austral spring, when stratospheric ozone ~~is known to~~ can influence the flow (Thompson and Solomon, 2002).

The use of time scale calculations based on individual decades allows us to provide a rough estimate ~~on the uncertainty associated with these time scales by computing~~ of the sampling uncertainty of τ. We quantify the sampling uncertainty by the standard deviation ~~between decades~~of the τ values obtained separately for each decade. In the Southern Hemisphere we can only trust the satellite era, giving us only three samples, the minimum needed to estimate the variance. In the Northern Hemisphere, however, we can take advantage of reanalyses since 1958, giving us 5 full decades. The sampling uncertainty is substantial. We quantify it in relative terms in Fig. 12c,d, which allows us to effectively compare both tropospheric and stratospheric values in one plot.

For the NAM, relative uncertainty in the time scale τ is between 10 and 20% over most of the annual cycle, and approaches 40% in the lower stratosphere during summer (a time, however, when the annular mode variability is quite weak). The longer time scales in the stratospheric winter, however, are comparatively robust. For the SAM (where the three decades give us only a rudimentary estimate of uncertainty), the tropospheric time scale – notably the marked increase in November and December – still appears statistically robust. There is great uncertainty in the early spring values of τ in the stratosphere, and τ is poorly constrained in austral spring (February-March) throughout the atmosphere.

We contrast the sampling uncertainty in τ with the spread associated with different reanalyses in the lower panels. As with the vortex event composites, we quantify the reanalysis uncertainty by the standard deviation of τ values across the four modern reanalyses (averaged for three separate decadal calculations for which all reanalyses are available). The reanalyses agree very well with each other, with a relative uncertainty less than 0.05 at most pressure levels and times during the year. There is some disagreement on time scales in the boreal stratosphere during summer (when variability is weak, see also Fig. 4) and in austral stratosphere during winter. This latter case is the only instance where the reanalysis uncertainty becomes comparable to the sampling uncertainty.

The fact that sampling uncertainty generally dwarfs the reanalysis uncertainty indicates that the main limiting factor is the length of the observational record. ~~Two additional decades of high~~ High quality reanalysis appear to be possible in the boreal hemisphere ~~, based in~~ from 1958-1978 based on our analysis in Sections 6, giving us the potential to increase the full record from approximately 40 years to 60. ~~A~~ As discussed in the context of vortex events, a 50% increase in sample size reduces the uncertainty in the mean ~~by $\sqrt{1/1.5}$, or about 80~~time scale by about 20%.

9 Conclusions

We have shown that the annular modes are very ~~well represented in the modern , second generation~~ consistently represented by the modern reanalyses. ERA-I, JRA-55, MERRA2, and ~~NOAA-CFSR~~ CFSR/CFSv2 can be used interchangeably for the purposes of quantifying the annular modes of our atmosphere. Preliminary analysis of ERA5 suggests equally high skill in this reanalysis as well. As detailed in Section 5, however, the other reanalyses are also quite good during the satellite era, albeit demonstrably less consistent in the austral hemisphere.

As discussed in Section 6, reanalysis of the Northern Hemisphere from 1958-1979 appears to be very consistent. The fact that the JRA-55C reanalysis (which explicitly excludes satellite observations) can match the full reanalyses in recent decades suggests that conventional observations are sufficient for a highly accurate representation of the Northern Annular Mode. In contrast, there is very little consistency in the representation of the Southern Annular Mode before satellite observations were available. In fact, near the surface, where one can estimate SAM variability on monthly time scales using station data (Marshall, 2003), the surface-input reanalyses (ERA-20C, NOAA-20CR v2 and v2c) appear superior to the full-input reanalyses. These surface-input reanalyses exhibit nearly uniform correlation with the station-based index in the pre- and post-satellite periods, but we have little means of evaluating their representation of stratospheric variability before 1979.

In Section 8, we highlighted the fact that our ability to quantify the downward coupling and temporal variability of the annular modes is limited primarily by sampling uncertainty. This suggests that two additional decades of reanalysis have substantial value for quantifying the variability of the boreal jet stream and polar vortex. This is important for both understanding the coupling between eddies and the mean flow (e.g., Lorenz and Hartmann, 2001) and testing the fidelity of atmospheric models (e.g., Gerber et al., 2010). We encourage reanalysis centers to consider extending their modern products back until 1958 or earlier.

Finally, we have documented the ability of reanalyses constrained only by surface observations to capture variability of the annular modes in the free troposphere and stratosphere. Both NOAA-20CR and ERA-20C accurately capture annular mode variability up to the extratropical tropopause. NOAA-20CR, both versions 2 and 2c, rapidly looses the ability to track the annular modes above the tropopause. ERA-20C, however, appears to capture much of the variability high into the stratosphere, notably capturing more than 50% of the variance as high as 10 hPa in both hemispheres.

The ability of ~~surface-only~~ surface-input reanalyses to constrain the stratosphere reflects the tight coupling between the troposphere and stratosphere. It should not be interpreted in a causal sense, i.e., that half of stratospheric variability is caused by surface observations. The issue of causality is important in that a number of studies have shown that modifying the stratosphere (for example, by nudging it toward observations or toward its climatology) can influence the tropospheric flow. While nudging is not equivalent to data assimilation, caution should be exercised in interpreting these experiments as one layer of the atmosphere driving another.

Data availability. The geopotential height data used to compute the annular modes is publicly available at the Centre for Environmental Data Analysis (Martineau, 2017) and documented in Martineau et al. (2018b). MERRA (Global Modeling and Assimilation Office, 2008) and MERRA2 (Global Modeling and Assimilation Office, 2015) output were obtained in the production of this data set.

Competing interests. There are no competing interest present.

Acknowledgements. EPG acknowledges support from the US National Science Foundation through grant AGS-1546585 to New York University. ~~We thank the~~ PM acknowledges support as an international research fellow of the Japan Society for the Promotion of Science. We thank Julie Arblaster and two anonymous reviewers for helpful comments on earlier versions of the manuscript, the reanalysis centers for making their data publicly available, and the Statosphere-troposphere Processes And their Role in Climate (SPARC) activity of the World Climate Research Program for their help in organizing the SPARC-Reanalysis Intercomparison Project, which inspired this research.

References

Ambaum, M. H. P., Hoskins, B. J., and Stephenson, D. B.: Arctic Oscillation or North Atlantic Oscillation?, J. Climate, 14, 3495–3507, 2001.

Baldwin, M. P. and Dunkerton, T. J.: Stratospheric Harbingers of Anomalous Weather Regimes, Science, 294, 581–584, 2001.

Baldwin, M. P. and Thompson, D. W. J.: A critical comparison of stratosphere-troposphere coupling indices, Quart. J. Roy. Meteor. Soc., 135, 1661–1672, 2009.

Baldwin, M. P., Stephenson, D. B., Thompson, D. W. J., Dunkerton, T. J., Charlton, A. J., and O'Neill, A.: Stratospheric Memory and Skill of Extended-Range Weather Forecasts, Science, 301, 636–640, 2003.

Barnes, E. A. and Hartmann, D. L.: Testing a theory for the effect of latitude on the persistence of eddy-driven jets using CMIP3 simulations, Geophys. Res. Lett., 37, L15 801, https://doi.org/10.1029/2010GL044144, 2010.

Barnes, E. A., Hartmann, D. L., Frierson, D. M. W., and Kidston, J.: Effect of latitude on the persistence of eddy-driven jets, Geophys. Res. Lett., 37, L11 804, https://doi.org/10.1029/2010GL043199, 2010.

Barsugli, J. J. and Battisti, D. S.: The basic effects of atmosphere–ocean thermal coupling on midlatitude variability, J. Atmos. Sci., 55, 477–493, 1998.

Black, R. X., McDaniel, B. A., and Robinson, W. A.: Stratosphere-Troposphere Coupling during Spring Onset, J. Climate, 19, 4892–4901, 2006.

Bosilovich, M., Akella, S., Coy, L., Cullather, R., Draper, C., Gelaro, R., Kovach, R., Liu, Q., Molod, A., Norris, P., Wargan, K., Chao, W., Reichle, R., Takacs, L., Vikhliaev, Y., Bloom, S., Collow, A., Firth, S., Labow, G., Partyka, G., Pawson, S., Reale, O., Schubert, S. D., and Suarez, M.: MERRA-2: Initial evaluation of the climate, NASA Tech. Rep. Series on Global Modeling and Data Assimilation, NASA/TM–2015-104606, 43, 2015.

Butler, A. H., Sjoberg, J. P., Seidel, D. J., and Rosenlof, K. H.: A sudden stratospheric warming compendium, Earth Syst. Sci. Data, 9, 63–76, https://doi.org/10.5194/essd-9-63-2017, 2017.

Byrne, N. J., Shepherd, T. G., Woollings, T., and Plumb, R. A.: Annular modes and apparent eddy feedbacks in the Southern Hemisphere, Geophys. Res. Lett., 43, 3897–3902, https://doi.org/10.1002/2016GL068851, 2016.

Charlton-Perez, A. J., Baldwin, M. P., Birner, T., Black, R. X., Butler, A. H., Calvo, N., Davis, N. A., Gerber, E. P., Gillet, N., Hardiman, S., Kim, J., Krueger, K., Lee, Y.-Y., Manzini, E., McDonald, B. A., Polvani, L. M., Reichler, T., Shaw, T. A., Sigmond, M., Son, S.-W., Tohey, M., Wilcox, L., Yoden, S., Christiansen, B., Lott, F., Shindell, D., Yukimoto, S., and Watanabe, S.: On the lack of stratospheric dynamical variability in low-top versions of the CMIP5 models, J. Geophys. Res., 118, 2494–2505, https://doi.org/10.1002/jgrd.50125, 2013.

Compo, G. P., Whitaker, J. S., Sardeshmukh, P. D., Matsui, N., Allan, R. J., Yin, X., Gleason, B. E., Vose, R. S., Rutledge, G., Bessemoulin, P., Brönnimann, S., Brunet, M., Crouthamel, R. I., Grant, A. N., Groisman, P. Y., Jones, P. D., Kruk, M. C., Kruger, A. C., Marshall, G. J., Maugeri, M., Mok, H. Y., Nordli, Ø., Ross, T. F., Trigo, R. M., Wang, X. L., Woodruff, S. D., and Worley, S. J.: The Twentieth Century Reanalysis Project, Quart. J. Roy. Meteor. Soc., 137, 1–28, https://doi.org/10.1002/qj.776, 2011.

Compo, G. P., Whitaker, J., Sardeshmukh, P., Giese, B., and Brohan, P.: Intercomparison of an improved 20th Century Reanalysis version "2c" (1850-2012), American Meteorological Society Annual meeting, Phoenix, AZ, 2015.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Mor-

crette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., and Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, Quart. J. Roy. Meteor. Soc., 137, 553–597, https://doi.org/10.1002/qj.828, 2011.

Deser, C.: On the teleconnectivity of the "Arctic Oscillation", Geophysical Research Letters, 27, 779–782, https://doi.org/10.1029/1999GL010945, 2000.

Feldstein, S. B. and Lee, S.: Is the Atmospheric Zonal Index Driven by an Eddy Feedback?, J. Atmos. Sci., 55, 3077–3086, 1998.

Fujiwara, M., Wright, J. S., Manney, G. L., Gray, L. J., Anstey, J., Birner, T., Davis, S., Gerber, E. P., Harvey, V. L., Hegglin, M. I., Homeyer, C. R., Knox, J. A., Krüger, K., Lambert, A., Long, C. S., Martineau, P., Molod, A., Monge-Sanz, B. M., Santee, M. L., Tegtmeier, S., Chabrillat, S., Tan, D. G. H., Jackson, D. R., Polavarapu, S., Compo, G. P., Dragani, R., Ebisuzaki, W., Harada, Y., Kobayashi, C., McCarty, W., Onogi, K., Pawson, S., Simmons, A., Wargan, K., Whitaker, J. S., and Zou, C.-Z.: Introduction to the SPARC Reanalysis Intercomparison Project (S-RIP) and overview of the reanalysis systems, Atmos. Chem. Phys., 17, 1417–1452, https://doi.org/10.5194/acp-17-1417-2017, 2017.

Garfinkel, C. I., Waugh, D. W., and Gerber, E. P.: The Effect of Tropospheric Jet Latitude on Coupling between the Stratospheric Polar Vortex and the Troposphere, J. Climate, 26, 2077–2095, https://doi.org/10.1175/JCLI-D-12-00301.1, 2013.

Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., Randles, C. A., Darmenov, A., Bosilovich, M. G., Reichle, R., Wargan, K., Coy, L., Cullather, R., Draper, C., Akella, S., Buchard, V., Conaty, A., da Silva, A. M., Gu, W., Kim, G.-K., Koster, R., Lucchesi, R., Merkova, D., Nielsen, J. E., Partyka, G., Pawson, S., Putman, W., Rienecker, M., Schubert, S. D., Sienkiewicz, M., and Zhao, B.: The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2), J. Climate, 30, 5419–5454, https://doi.org/10.1175/JCLI-D-16-0758.1, 2017.

Gerber, E. P. and Thompson, D. W. J.: What Makes an Annular Mode “Annular”?, J. Atmos. Sci., 74, 317–332, https://doi.org/10.1175/JAS-D-16-0191.1, 2017.

Gerber, E. P. and Vallis, G. K.: A Stochastic Model for the Spatial Structure of Annular Patterns of Variability and the NAO, J. Climate, 18, 2102–2118, https://doi.org/10.1175/JCLI3337.1, 2005.

Gerber, E. P. and Vallis, G. K.: Eddy-Zonal Flow Interactions and the Persistence of the Zonal Index, J. Atmos. Sci., 64, 3296–3311, https://doi.org/10.1175/JAS4006.1, 2007.

Gerber, E. P., Polvani, L. M., and Ancukiewicz, D.: Annular Mode Time Scales in the Integovernmental Panel on Climate Change Fourth Assessment Report Models, Geophys. Res. Lett., 35, L22 707, https://doi.org/10.1029/2008GL035712, 2008a.

Gerber, E. P., Voronin, S., and Polvani, L. M.: Testing the Annular Mode Autocorrelation Timescale in Simple Atmospheric General Circulation Models, Mon. Wea. Rev., 136, 1523–1536, https://doi.org/10.1175/2007MWR2211.1, 2008b.

Gerber, E. P., Baldwin, M. P., Akiyoshi, H., Austin, J., Bekki, S., Braesicke, P., Butchart, N., Chipperfield, M., Dameris, M., Dhomse, S., Frith, S. M., Garcia, R. R., Garny, H., Gettelman, A., Hardiman, S. C., Karpechko, A., Marchand, M., Morgenstern, O., Nielsen, J. E., Pawson, S., Peter, T., Plummer, D. A., Pyle, J. A., Rozanov, E., Scinocca, J. F., Shepherd, T. G., and Smale, D.: Stratosphere-Troposphere Coupling and Annular Mode Variability in Chemistry-Climate Models, J. Geophys. Res., 115, D00M06, https://doi.org/10.1029/2009JD013770, 2010.

Global Modeling and Assimilation Office: inst3_3d_asm_Cp: MERRA 3D IAU State, Meteorology Instantaneous 3-hourly V5.2.0, https://doi.org/10.5067/8D4LU4390C4S, http://disc.sci.gsfc.nasa.gov/mdisc/, Goddard Earth Sciences Data and Information Services Center (GES DISC), Accessed: 2017-10-4, 2008.

Global Modeling and Assimilation Office: MERRA-2 inst3_3d_asm_Np: 3d, 3-Hourly, Instantaneous, Pressure-Level, Assimilation, Assimilated Meteorological Fields V5.12.4, https://doi.org/10.5067/QBZ6MG944HW0, http://disc.sci.gsfc.nasa.gov/mdisc/, Goddard Earth Sciences Data and Information Services Center (GES DISC), Accessed: 2017-07-5, 2015.

Kalnay, E., Kanamitsua, M., Kistler, R., Collinsa, W., Deavena, D., Gandina, L., Iredella, M., Sahaa, S., Whitea, G., Woollena, J., Zhua, Y., Leetmaaa, A., Reynoldsa, R., Chelliahb, M., Ebisuzakib, W., Higginsb, W., Janowiakb, J., Mob, K., Ropelewskib, C., Wangb, J., Jennec, R., and Josephc, D.: The NCEP/NCAR 40-year reanalysis project, Bull. Am. Meteor. Soc., 77, 437–471, 1996.

Kanamitsu, M., Ebisuzaki, W., Woollen, J., Yang, S.-K., Hnilo, J. J., Fiorino, M., and Potter, G. L.: NCEP–DOE AMIP-II reanalysis (R-2), Bull. Am. Meteor. Soc., 83, 1631–1643, 2002.

Keeley, S. P. E., Sutton, R. T., and Shaffrey, L. C.: Does the North Atlantic Oscillation show unusual persistence on intraseasonal timescales?, Geophys. Res. Lett., 36, L22 706, https://doi.org/10.1029/2009GL040367, 2009.

Kidston, J. and Gerber, E. P.: Intermodel Variability of the Poleward Shift of the Austral Jet Stream in the CMIP3 Integrations Linked to Biases in 20th Century Climatology, Geophys. Res. Lett., 37, L09 708, https://doi.org/10.1029/2010GL042873, 2010.

Kistler, R.and Collins, W., Saha, S., G., W., Woollen, J., Kalnay, E., Chelliah, M., Ebisuzaki, W., Kanamitsu, M., Kousky, V.and van den Dool, H., Jenne, R., and Fiorino, M.: The NCEP–NCAR 50-year reanalysis: Monthly means CD-ROM and documentation, Bull. Am. Meteor. Soc., 82, 247–267, 2001.

Kobayashi, C., Endo, H., Ota, Y., Kobayashi, S., Onoda, H., Harada, Y., Onogi, K., and Kamahori, H.: Preliminary results of the JRA-55C, an atmospheric reanalysis assimilating conventional observations only, Sci. Online Lett. Atmos., 10, 78–82, https://doi.org/10.2151/sola.2014-016, 2014.

Kobayashi, S., Ota, Y., Harada, Y., Ebita, A., Moriya, M., Onoda, H., Onogi, K., Kamahori, H., Kobayashi, C., Endo, H., Miyaoka, K., and Takahashi, K.: The JRA-55 reanalysis: General specifications and basic characteristics, J. Meteorol. Soc. Jpn., 93, 5–48, https://doi.org/10.2151/jmsj.2015-001, 2015.

L'Heureux, M. L. and Thompson, D. W. J.: Observed Relationships between the El Niño–Southern Oscillation and the Extratropical Zonal-Mean Circulation, J. Climate, 19, 279–287, 2006.

Limpasuvan, V. and Hartmann, D. L.: Wave-Maintained Annular Modes of Climate Variability., J. Climate, 13, 4414–4429, 2000.

Long, C. S., Fujiwara, M., Davis, S., Mitchell, D. M., and Wright, C. J.: Climatology and interannual variability of dynamic variables in multiple reanalyses evaluated by the SPARC Reanalysis Intercomparison Project (S-RIP), Atmos. Chem. Phys., 17, 14 593–14 629, https://doi.org/10.5194/acp-17-14593-2017, 2017.

Lorenz, D. J. and Hartmann, D. L.: Eddy-zonal flow feedback in the Southern Hemisphere, J. Atmos. Sci., 58, 3312–3327, 2001.

Marshall, G. J.: Trends in the Southern Annular Mode from Observations and Reanalyses, J. Climate, 16, 4134–4143, 2003.

Martineau, P.: S-RIP: Zonal-mean dynamical variables of global atmospheric reanalyses on pressure levels, Centre for Environmental Data Analysis, May 2018, https://doi.org/10.5285/b241a7f536a244749662360bd7839312, 2017.

Martineau, P., Son, S.-W., Taguchi, M., and Butler, A. H.: A comparison of the momentum budget in reanalysis datasets during sudden stratospheric warming events, Atmos. Chem. Phys., 18, 7169–7187, https://doi.org/10.5194/acp-18-7169-2018, 2018a.

Martineau, P., Wright, J. S., Zhu, N., and Fujiwara, M.: Zonal-mean data set of global atmospheric reanalyses on pressure levels, Earth Syst. Sci. Data, 10, 1925–1941, https://doi.org/10.5194/essd-10-1925-2018, 2018b.

Namias, J.: The Index Cycle and its Role in the General Circulation, J. Meteorology, 7, 130–139, 1950.

Onogi, K., Tsutsui, J., Koide, H., Sakamoto, M., Kobayashi, S., Hatsushika, H., Matsumoto, T., Yamazaki, N., Kamahori, H., Takahashi, K., Kadokura, S., Wada, K., Kato, K., Oyama, R., Ose, T., Mannoji, N., and Taira, R.: The JRA-25 reanalysis, J. Meteorol. Soc. Jpn., 85, 369–432, https://doi.org/10.2151/jmsj.85.369, 2007.

Osprey, S. and Ambaum, M. H. P.: Evidence for the chaotic origin of Northern Annular Mode variability, Geophys. Res. Lett., 38, https://doi.org/10.1029/2011GL048181, 2011.

Poli, P., Hersbach, H., Tan, D., Dee, D. P., Berrisford, P., Simmons, A., Vitart, F., Laloyaux, P., Tan, D. H., Peubey, C., Thépaut, J.-N., Trémolet, Y., Hólm, E., Bonavita, M., Isaksen, L., and Fisher, M.: ERA-20C: An atmospheric reanalysis of the 20th century, J. Climate, 29, 4083–4097, https://doi.org/10.1175/JCLI-D-15-0556.1, 2016.

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

Robinson, W. A.: Does Eddy Feedback Sustain Variability in the Zonal Index, J. Atmos. Sci., 53, 3556–3569, 1996.

Rossby, C. G.: Relations between Variation in the Intesity of the Zonal Circulation and the Displacements of the Semi-permanent Centers of Action, J. Marine Res., 2, 38–55, 1939.

Saha, S., Moorthi, S., Pan, H.-L., Wu, X., Wang, J., Nadiga, S., Tripp, P., Kistler, R., Woollen, J., Behringer, D., Liu, H., Stokes, D., Grumbine, R., Gayno, G., Wang, J., Hou, Y.-T., Chuang, H.-Y., Juang, H.-M. H., Sela, J., Iredell, M., Treadon, R., Kleist, D., Delst, P. V., Keyser, D., Derber, J., Ek, M., Meng, J., Wei, H., Yang, R., Lord, S., van den Dool, H., Kumar, A., Wang, W., Long, C., Chelliah, M., Xue, Y., Huang, B., Schemm, J.-K., Ebisuzaki, W., Lin, R., Xie, P., Chen, M., Zhou, S., Higgins, W., Zou, C.-Z., Liu, Q., Chen, Y., Han, Y., Cucurull, L., Reynolds, R. W., Rutledge, G., and Goldberg, M.: The NCEP climate forecast system reanalysis, Bull. Am. Meteor. Soc., 91, 1015–1057, https://doi.org/10.1175/2010BAMS3001.1, 2010.

Saha, S., Moorthi, S., Wu, X., Wang, J., Nadiga, S., Tripp, P., Behringer, D., Hou, Y.-T., ya Chuang, H., Iredell, M., Ek, M., Meng, J., Yang, R., Mendez, M. P., van den Dool, H., Zhang, Q., Wang, W., Chen, M., and Becker, E.: The NCEP Climate Forecast System Version 2, J. Climate, 27, 2185–2208, https://doi.org/10.1175/JCLI-D-12-00823.1, 2014.

Sheshadri, A. and Plumb, R. A.: Propagating Annular Modes: Empirical Orthogonal Functions, Principal Oscillation Patterns, and Time Scales, J. Atmos. Sci., 74, 1345–1361, https://doi.org/10.1175/JAS-D-16-0291.1, 2017.

Sigmond, M., Scinocca, J. F., Kharin, V. V., and Shepherd, T. G.: Enhanced seasonal forecast skill following stratospheric sudden warmings, Nature Geosci., https://doi.org/10.1038/NGEO1698, 2013.

Simpson, I. R. and Polvani, L. M.: Revisiting the relationship between jet position, forced response, and annular mode variability in the southern midlatitudes, Geophys. Res. Lett., 43, 2896–2903, https://doi.org/10.1002/2016GL067989, 2016.

Son, S.-W., Gerber, E. P., Perlwitz, J., Polvani, L. M., Gillet, N., Seo, K.-Y., Eyring, V., Shepherd, T. G., Waugh, D., Akiyoshi, H., Austin, J., Baumgaertner, A., Bekki, S., Braesicke, P., Brühl, C., Butchart, N., Chipperfeld, M. P., Cugnet, D., Dameris, M., Dhomse, S., Frith, S., Garny, H., Garcia, R., Hardiman, S. C., Jöckel, P., Lamarque, J. F., Mancini, E., Marchand, M., Michou, M., Nakamura, T., Morgenstern, O., Pitari, G., Plummer, D. A., Pyle, J., andJ. F. Scinocca, E. R., Shibata, K., Smale, D., Teyssèdre, H., Tian, W., and Yamashita, Y.: The Impact of Stratospheric Ozone on Southern Hemisphere Circulation Change: A Multimodel Assessment, J. Geophys. Res., 115, D00M07, https://doi.org/10.1029/2010JD014271, 2010.

Thompson, D. W. J. and Solomon, S.: Interpretation of recent Southern Hemisphere Climate Change, Science, 296, 895–899, 2002.

Thompson, D. W. J. and Wallace, J. M.: The Arctic Oscillation signature in the wintertime geopotential height and temperature fields., Geophys. Res. Lett., 25, 1297–1300, 1998.

Thompson, D. W. J. and Wallace, J. M.: Annular modes in the extratropical circulation. Part I: Month-to-month variability., J. Climate, 13, 1000–1016, 2000.

Uppala, S. M., Kållberg, P. W., Simmons, A. J., Andrae, U., Bechtold, V. D., Fiorino, M., Gibson, J. K., Haseler, J., Hernandez, A., Kelly, G. A., Li, X., Onogi, K., Saarinen, S., Sokka, N., Allan, R. P., Andersson, E., Arpe, K., Balmaseda, M. A., Beljaars, A. C. M., Van De Berg, L., Bidlot, J., Bormann, N., Caires, S., Chevallier, F., Dethof, A., Dragosavac, M., Fisher, M., Fuentes, M., Hagemann, S., Holm, E., Hoskins, B. J., Isaksen, L., Janssen, P. A. E. M., Jenne, R., McNally, A. P., Mahfouf, J. F., Morcrette, J. J., Rayner, N. A., Saunders, R. W., Simon, P., Sterl, A., Trenberth, K. E., Untch, A., Vasiljevic, D., Viterbo, P., and Woollen, J.: The ERA-40 Reanalysis, Quart. J. Roy. Meteor. Soc., 131, 2961–3012, 2005.

Vallis, G. K., Gerber, E. P., Kushner, P. J., and Cash, B. A.: A Mechanism and Simple Dynamical Model of the North Atlantic Oscillation and Annular Modes., J. Atmos. Sci., 61, 264–280, 2004.

Zurita-Gotor, P., Blanco-Fuentes, J., and Gerber, E. P.: The Impact of Baroclinic Eddy Feedback on the Persistence of Jet Variability in the Two-Layer Model, J. Atmos. Sci., 71, 410–429, https://doi.org/10.1175/JAS-D-13-0102.1, 2014.

Figure 1. To illustrate our simplified method for calculating the annular mode indices, (a) shows anomalous zonal mean geopotential height, $\overline{Z'}$, at 10 hPa for one year, 2007. As detailed in the text, anomalies are defined relative to a smoothed annual cycle. Panel (b) shows global and polar cap averages of Z' for this period, and (c) the Northern and Southern Annular Mode indices at this level, defined as -1$\times$ the respective polar cap averages less the global mean, and then normalized to have unit variance. The choice of year 2007 was arbitrary, but did exhibit a Sudden Stratospheric Warming on 24 February, which is associated with high geopotential height over the polar cap and low NAM index around day 55.

Figure 2. ~~Consistency of~~ Example annular mode patterns, computed over the period 1981-2010 from ERA-Interim (~~a~~solid lines) ~~the Northern~~ and ERA-20C (~~b~~dashed lines) ~~Southern Annular Mode indices based on~~ at three select levels. As the ~~modern reanalyses~~geopotential height anomalies increase at higher levels, ~~as a function~~ the amplitude of ~~pressure~~the 100 and 10 hPa patterns has been divided by 2 and 6, respectively. The ~~pairwise correlation coefficient (R^2) between ERA-I, JRA-55, MERRA2, and NCEP-CFSR~~ 1000 hPa patterns are ~~plotted~~nearly identical to the 850 hPa; we show the ~~dashed colors indicating~~ latter to minimize the ~~respective pairs, e.g., green-red dashing indicates~~ need for extrapolation in the ~~square correlation between MERRA2 and NCEP-CFSR~~austral hemisphere.

Figure 3. Consistency of (a) the Northern and (b) Southern Annular Mode indices in the modern reanalyses over the period 1981-2010, quantified by the pairwise correlation coefficient (R^2) between the annular mode indices computed from ERA-Interim, JRA-55, MERRA2, and CFSR, plotted as a function of height. The dashed colors indicate the respective indices that are being compared; for example, the green-red dashing indicates the square correlation between MERRA2 and CFSR. The square of the correlation coefficient, R^2, indicates the fraction of variance shared between two time series. Hence one may conclude that the NAM indices in MERRA2 and CFSR share greater than 96% of the variance at all levels.

Figure 4. Consistency of (a) the Northern and (b) Southern Annular Mode indices in the modern reanalyses, as a function of pressure and time of year, for the period ~~1980~~ 1981 to ~~2016. The shading indicates~~ 2010. Consistency was quantified by the average pairwise R^2 correlation between ERA-I, JRA-55, MERRA2, and ~~NCEP-CFSR~~CFSR, and indicated by the color shading. Contours indicate the variance of the REM annular mode, which has been normalized to have unit variance in the annual mean at each level; the interval is 0.5 units of variance, with an additional dashed contour at 0.1.

Figure 5. ~~The squared correlation between~~ Comparison of the ~~(a,c) Northern~~ SAM in CFSR and ~~(b~~CFSv2 with the other modern reanalyses. As in Fig. 3, ~~d) Southern Annular Mode indices computed~~ the R^2 correlation is plotted: solid lines show a comparison with CFSR over the period 2005-2010, while dashed lines show a comparison with CFSv2 over the period 2011-2016. Preliminary output from ~~each individual~~ the ERA5 reanalysis ~~with~~ is available starting in 2008, and could be included in the ~~Reanalyses Ensemble Mean (REM)~~latter comparison. We ~~use~~ show only R^2 correlation with the ~~standard WMO three decade climatological period~~SAM indices, ~~1981-2010, for all reanalyses except ERA-40, where analysis is based on 1981-2001~~as there were no discernable differences with the NAM indices between these two periods. ~~Panels~~ Panel (b~~,d~~) ~~show~~ shows the same ~~data~~as (a), but ~~zoomed in to highlight~~ we have expanded the ~~more accurate reanalyses. Note the difference in the~~ range of the ~~abscissa: the NAM is very well constrained in all reanalyses that assimilate free atmospheric observations~~x-axis to focus on mid-levels.

Figure 6. The squared correlation between the (a,b) Northern and (c,d) Southern Annular Mode indices computed from each individual reanalysis with the Reanalyses Ensemble Mean (REM). We use the standard WMO three decade climatological period, 1981-2010, for all reanalyses except ERA-40, where analysis is based on 1981-2001. Panels (b,d) show the same data, but zoomed in to highlight the more accurate reanalyses. Note the difference in the range of the abscissa: the NAM is very well constrained in all reanalyses that assimilate free atmospheric observations.

Figure 7. The squared correlation between (top) Northern and (bottom) Southern Annular Mode indices in various reanalyses with (left) JRA-55 and (right) NCEP-R1 during the pre-satellite era, 1958-1978.

Figure 8. Comparison between the Northern Annular Mode timeseries from the Reanalysis Ensemble Mean (REM) and ERA-20C at 10, 100, 1000 hPa over a two year period, 1 July 2006 to 1 July 2008. Yellow dots indicate the occurrence of two Sudden Stratospheric Warming events during this period.

Figure 9. Correlation of the (top) NAM and (bottom) SAM computed from ERA-20C, as a function of time, with JRA-55, NCEP-R1, and the NOAA 20C reanalyses. The correlation is computed over a moving 11 year window, centered about the time given on the x-axis, i.e., the values at 1980 correspond to correlation between 1975 and 1985. ERA-40 was not included, as it provides comparable information to JRA-55 and NCEP-R1.

Figure 10. Composites of the Northern Annular Mode indices as a function of lag and pressure for (a) weak and (b) strong vortex events, based on JRA-55 reanalyses over the period 1958-2016. Following Baldwin and Dunkerton (2001), weak (strong) events are identified when the NAM index at 10 hPa drops below -3 (rises above 1.5), and must be separated by a minimum of 30 days. Panels (c) and (d) show the event-to-event standard deviation for weak and strong vortex events, respectively.

Figure 11. Three estimates of the uncertainty in the NAM index evolution around weak and strong events. (a) and (b) show the *sampling uncertainty* in the mean weak/strong composites shown in Fig. 10a,b, expressed as a 1 standard deviation error bound. Panels (c) and (d) show a first estimate of the *reanalysis uncertainty*: the standard deviation between composites of weak/strong vortex composites based on the 4 modern reanalyses (ERA-I, JRA-55, MERRA2, and ~~NOAA-CFSR~~CFSR/CFSv2) over the period 1980-2016, where events are determined independently in each reanalysis. Panels (e) and (f) show a refined estimate of the *reanalysis uncertainty*: the standard deviation of weak/strong vortex composites based on the 4 modern reanalyses, but now using a standardized set of event dates. See the text for further details.

Figure 12. The annual cycle in the time scale of annular mode variability, τ, and relative uncertainty associated with the finite record and spread between the reanalyses. As detailed in the text, the top panels show the decorrelation time scale τ of the (a) Northern and (b) Southern Annular Modes in JRA-55 analysis, based on the average of 5 and 3 independent decadal calculations (1960-9, 1970-9, ..., 2000-2009 for the NH, only the last three decades for the SH). The middle panels show the sampling uncertainty of τ for (c) the NAM and (d) the SAM: the standard deviation of the decadal calculations divided by the square root of the sample size (5 and 3, respectively). We show the uncertainty in a relative sense, normalizing by the mean time scale shown in the upper panels; otherwise errors in the troposphere (where τ is smaller) appear insignificant. The bottom panels show the reanalysis uncertainty in τ for the (e) NAM and (f) SAM: the standard deviation of τ between calculations using ERA-I, JRA-55, MERRA2, and ~~NOAA-CFSR~~CFSR. The standard deviation was computed across three separate decadal calculations (1980-9, 1990-9, and 2000-9) and then averaged.

Table 1. The atmospheric reanalyses analyzed in this study. Full-input reanalyses take into account new observations as they become available, while a conventional-input reanalysis specifically exclude satellite-based observations and surface-input reanalyses ingest only surface measurements. The availability of surface and conventional observations does fluctuate with time, however, such that these reanalyses are not immune to spurious trends associated with changes in the observation network. The JRA-55AMIP provides an free running integration of the JRA-55 atmospheric model forced with observed sea-surface temperatures, as in an Atmospheric Model Intercomparison Project (AMIP) integration.

Reanalysis System	Coverage	Reference
Full-input reanalyses		
ERA-40	9/1957-8/2002	Uppala et al. (2005)
ERA-Interim	1/1979-present	Dee et al. (2011)
~~ERA-20C~~ ERA5	~~Poli et al. (2016)~~2008-2016*	—
JRA-25/JCDAS (JRA-25)	1/1979-1/2014	Onogi et al. (2007)
JRA-55	~~Kobayashi et al. (2015)~~JRA-55C 1/1958-present	~~Kobayashi et al. (2014)~~JRA-55AMIP Kob
MERRA	1/1979-2/2016	Rienecker et al. (2011)
~~MERRA2~~ MERRA-2	~~Bosilovich et al. (2015)~~ 1/1980-present	Gelaro et al. (2017)
~~NCEP-NCAR R1~~ NCEP-R1	~~Kalnay et al. (1996)~~ 1/1948-present	Kalnay et al. (1996); Kistler et al. (2001)
~~NCEP-DOE R2~~ NCEP-R2	1/1979-present	Kanamitsu et al. (2002)
~~NCEP-CFSR*~~ CFSR	~~Saha et al. (2010, 2014)~~ 1/1979-12/2010	Saha et al. (2010)
~~NOAA-20CR~~ CFSv2	1/2011-present	Saha et al. (2014)
Conventional-input reanalyses		
JRA-55C	1/1979-12/2012	Kobayashi et al. (2014)
Surface-input reanalyses		
ERA-20C	1/1900-12/2010	Poli et al. (2016)
NOAA-CIRES 20CR v2 (20CR v2)	11/1869-12/2012	Compo et al. (2011)
~~NOAA-20CR~~ NOAA-CIRES 20CR v2c (20CR v2c)	1/1851-12/2011	Compo et al. (2015)
AMIP Simulation		
JRA-55AMIP	1/1958-12/2012	Kobayashi et al. (2015)

*ERA5 is currently in production, with the goal of making data available from 1950 onward.

Table 2. Correlation between the monthly mean Southern Annular Mode at 850 hPa with the Marshall (2003) station-based index during the pre- and post-satellite periods. Note that correlations are virtually identical if we compare with the 1000 hPa indices, as opposed to 850. While the station-based index is available from 1957 to present, we focus on these periods to optimize availability of reanalyses; in particular, ERA-40 is only available for complete years from 1958-2001.

reanalysis	1958-1978	1979-2001
ERA-40	0.51	0.87
ERA-20C	0.83	0.86
JRA-55	0.53	0.85
NCEP-R1	0.70	0.84
20CR v2	0.85	0.85
20CR v2c	0.84	0.86