I thank all three reviewers for their considered, constructive comments. They have significantly improved the manuscript. I have included my detailed responses in italics inline below.

Response to Reviewer 1

This paper presents an interesting and carefully presented study of the value of reanalyses covering years before 1979 for use in certain dynamical studies. It merits publication, but would benefit from minor revision to take the following comments into account.

(1) Title. As the study deals mainly with stratospheric dynamics, and stratospheric sudden warmings in particular, the author should consider changing the title so that dynamical studies becomes studies of stratospheric dynamics or studies of the dynamics of sudden stratospheric warmings. The study does not present much evidence concerning tropospheric dynamics, nor does it reference results on tropospheric dynamics from other studies

The focus is on stratosphere-troposphere coupling - Fig.5 through 10 all include some component of tropospheric dynamics, and one of the points made is that many of the open questions in this area revolve around tropospheric feedbacks (for which the momentum fluxes considered in the final figure play a central role). The suggested titles would thus mischaracterize the text. Moreover, the criteria developed for evaluating this period are just as applicable to other composite-based dynamical studies (for example); this was the justification for the title. Still, given the emphasis on stratosphere-troposphere coupling, I have added this to the title.

(2) Page 1, line 3. The word satellite is rightly in inverted commas in the abstract. But it needs to be made clear in the body of the paper that the satellite era begins before 1979, and that it is a simplification, albeit a reasonable one, to refer to the period up to 1978 as the radiosonde era, and the period from 1979 as the satellite era. In practice (and as discussed by Uppala et al. (2005) for those observations used in ERA40):

(i) The MSU and SSU sounding data that characterise the start of the satellite era are available from November 1978.

(ii) ERA40 and JRA55 assimilated radiances from the VTPR instrument available from late 1972 until early 1979.

(iii) ERA5 is currently assimilating BUV ozone data available from 1970. Ozone analyses provide implicit information on stratospheric dynamics.

(iv) Some cloudtracked wind data from satellites are available and used prior to 1979.

(v) Satellite imagery was used by the Australian Bureau of Meteorology to generate pseudo surfacepressure observations that were assimilated in ERA40 from 1972 to 1978, although these are not being used now in ERA5 to the best of my knowledge.

In addition, there are improvements over time to the observing system, as indeed discussed in the paper under review. It is perhaps worth noting in the paper that in the radiosonde era, and back to the late 1940s, soundings over the North Atlantic and (to a lesser extent) the North Pacific Ocean were provided from fixedposition weather ships that were retired once satellite soundings were shown to provide a sufficient alternative. This in part compensates for lack of satellite data in the earlier years for the northern hemisphere.

Thank you for these details. I have added further clarification of the use of 'satellite' and 'radiosonde' eras as convenient simplifications, emphasizing in the introduction both the availability of satellite data prior to 1979 as well as the availability of radiosonde data prior to 1958, and their continued importance after 1979. I have however avoided getting too much into the history of the observational network as I don't feel I have expertise or historical knowledge to do this justice.

(3) Page1, line 5. The word could should be avoided here. The paper demonstrates that the radiosonde era does extend the useful period of record back beyond 1979, so this should be made clear in the abstract. The sentence as it stands leaves the question still open.

This is a good point. The text has been changed as suggested.

(4) Page 1, Line 10. It is inappropriate to issue a blanket call for reanalysis centres to consider generating products prior to 1979. ECMWF did this for ERA40, and is currently producing analyses from 1950 onwards for ERA5. ECMWF has also studied use of radiosonde and other upperair data for the period 19391967 as reported in a paper by Hersbach et al. (2017, doi: 10.1002/qj.3040) that rather surprisingly is not referenced in the paper under review. JMA ran JRA55 from 1958 onwards, and will soon start production of JRA3Q, for which the plan is to start in the late 1940s. So these two major producers appear already to appreciate the value of products prior to 1979 though further evidence as provided by the paper under review is always welcome.

I have added citations and discussion of Hersbach et al. in several appropriate places (see also responses below), thank you for bringing this work to my attention. I have left in the recommendation that future reanalyses include this period as I see no reason not to do so. Of course reanalyses centers are responding to the needs of a huge range of users who benefit greatly from this service, and there are accordingly a wide diversity of priorities. My intent with this paper was to put on record a quantitative argument for why this period is of value, partly in the hope that it can be useful to reanalysis centers in justifying their use of resources.

(5) Page 2, line 8. It would be appropriate here to record that ECMWF is currently producing ERA5 reanalyses from 1950 onwards and that analyses from the late 1940s onwards are expected from JRA3Q.

This has been done.

(6) Page 2, line 18. A reference to Hersbach et al. (2017) could be introduced here.

I have introduced it a bit later in the introduction.

(7) Page 2, line 22. It could be referenced here that Simmons et al. (2005, J.Atmos. Sci, March) demonstrated that the ERA40 reanalysis was of sufficient quality in January 1958 to produce a good fiveday forecast of the splitvortex sudden warming that occurred during that month. Caveats were issued in this paper about the quality of the stratospheric analyses over the southern hemisphere prior to 1979, but these analyses nevertheless gave no indication of a major splitvortex event between 1957 and 1978 of the type observed in September 2002, a result consistent with analysis of the sparse radiosonde data available for the period.

A citation to Simmons et al. 2005 has been added here.

(8) Page 5, lines 12 to 14. The text here needs revising. It refers to uncertainties in observations but errors in forecast models and the assimilation process. In reality there are errors in observations, and uncertainties in modelling and assimilation due to the stochastic nature of some of the processes being dealt with. So one should not use one word for observations and another for models/assimilation.

I have reworded these and other sentences to avoid associating 'error' or 'uncertainty' specifically with the observations or the modelling/assimilation process.

(9) Page 5, equation (2). The upper limit of the second sum on the lefthand side of the equation should be N r not N s .

Changed - thank you for noticing this.

(10) Page 6, lines 21 and 22. Same comment as (8) regarding the use of the words observational uncertainty and errors in the forecast model and the assimilation process.

Reworded.

(11) Page 7, line 6. The text on Fig5 refers to sat and rad , whereas Fig 7 and the text refer to s and r . This should be rectified.

Fixed.

(12) Page 7, line 16. The lack of a strong balance constraint is a reasonable explanation for the reanalysis uncertainty in the tropical upper stratosphere. But reanalysis uncertainty is much lower than dynamical variability at 10hPa and below. This is presumably because radiosonde data alone are quite effective in constraining the QBO in the lower and middle stratosphere in reanalyses. A comment could be added to this effect.

This is a good point and has been commented on in the updated text.

(13) Page 7, lines 17 and 18. Manney et al. (2005) is the reference the author chooses to use here. But the deficiency of ERA40 under discussion was first identified in preparing a SPARC Report on stratospheric climatology, and this was published in a subsequent peerreviewed paper by Randel et al. (2004), i.e. earlier than the Manney et al. paper. So Randels paper would probably be a fairer reference. The problem was also acknowledged in Uppala et al.s (2005) writeup of ERA40.

The reference has been changed to Randel et al. 2004.

(14) Page 10, line 26. Delete the word at.

Done.

(15) Page 10, line 34. This is another place where a reference to Hersbach et al. (2017) could be added, as that paper discusses, inter alia, the utility of 1950s radiosonde data for analysing the QBO.

Agreed - this has been done.

(16) Page 11, line 23. following 1979 should at least be changed to following 1978 and more precisely could be written from late 1978 onwards. The subsequent reference to radiosondes being remarkably effective in constraining the boreal stratosphere from 1958 to 1978 perhaps can remain as is, in view of the results of JRA55C, even though VTPR data provide an additional constraint from late 1972. It perhaps should be recognised however that radiosonde observations continue to provide a constraint on the stratosphere from 1979 onwards. Satellite radiances (particularly from the TOVS instruments flown from 1978 until phased out between 1998 and 2006)have significant biases, and radiosondes play an important role in the biascorrection schemes for radiance data used by reanalysis centres, at least prior to the availability of substantial amounts of GPS radio occultation data from 2006 onwards. The betterquality reanalyses produced for the period from 1979 onwards is due to the combined use of radiosonde and satellite data, notwithstanding the labelling of the period as the satellite era.

This paragraph has been reworked to provide a better overview of the study, and to better reflect the presence of satellite data products from prior to 1979 as well as the continued value of radiosonde data.

(17) Page 12, line 15. be should be been.

Corrected.

(18) Page 12, line 21. Some rewording is required here, as it is a bit misleading to categorize the sudden warmings in fulldata reanalyses as a result of assimilated observations. They are a result of assimilating observations making use of a forecast model, and as such are a result of both forecastmodel dynamics and assimilated observations. The reanalyses that assimilate only surface observations demonstrate that assimilating upperair observations is important, but does not show that the forecast model is unimportant.

I agree with your underlying point, but I don't think its fair to infer what you suggest from the text. The sentence is comparing the respective roles of the forecast model and the observations in constraining the event dates in surface input reanalyses, stating that the former is more important than the latter. Nowhere does it say anything about full-input reanalyses; if you insisted on inferring the converse it would be that assimilated observations are more important than the forecast model, not that the forecast model is unimportant. They could just as well be of equal importance in full-input reanalyses.

(19) Page 12, lines 22 to 25. Newer reanalyses apply bias corrections to radiosonde data (generally following the work of Haimberger), and assimilating the biascorrected radiosonde data tends to control the biases of the reanalyses, at least in places and at levels radiosonde data are reasonably plentiful. No biascorrection of pre1979 radiosonde data was applied in ERA40, but the radiosonde data would nevertheless have limited systematic error in ERA40 to some extent. A change is not called for at this point in the paper, but consideration could be given to writing something earlier in the paper on this point.

While I appreciate this comment as this is exactly the kind of impprovement that should bring improved confidence in the representation of 'radiosonde' era circulation, it is difficult for me to see how to explicitly tie this bias-correction to systematic errors in general (that, for instance, may not occur where observationns are directly available). It wasn't clear to me where to add this point to the text.

(20) Page 12, lines 29 and 30. Comment (4) above, concerning the final sentence of the abstract, applies equally to this final sentence of section 6.

I have mentioned ERA-5 and JRA-3Q here.

Response to Reviewer 2

General Comments

The author presents a detailed comparative analysis of the quality of reanalyses data prior to 1979 and their potential inclusion in dynamical studies. In particular, he focuses on the analysis of relevant fields for the stratosphere-troposphere coupling. The results indicate that reanalysis data in the pre-satellite era is of sufficiently high quality to be considered together with the data of subsequent decades in these dynamical studies.

The manuscript is well written and the topic is certainly interesting for the scientific community, particularly that focused on stratosphere-troposphere coupling. The methodology applied for evaluating the quality of reanalysis data is also very thorough. However, in some cases I find the text a little bit dense particularly when describing Figures 6 and 7 and Section 4 and it would be advantageous for the manuscript to try to simplify that description. Thus, I recommend the publication of the manuscript after the mentioned minor correction and some other slight changes indicated below.

In light of this and other reviewer comments I have spent some time trying to make this discussion clearer and more straightforward.

Specific Comments

Page 5 Lines 5-8: The author indicates that the shift of the seasonal peak of SSWs in the satellite era with respect to the whole period (1958-2010) is only just due to the consideration of a longer database. I think the author could discuss a little bit more about this. Otherwise, the reader might get the impression that this is only a possible bias due to the lack of assimilated satellite data in the pre-satellite period. In contrast, it could be also related to multidecadal climate variability. Indeed, there are some studies that have also shown a change in the seasonality of SSWs in model simulations (e.g.: Ayarzagena et al. 2013). Finally, I would recommend citing here Gmez-Escolar et al. (2012) that already showed the change in the seasonal distribution of SSWs between the pre- and post-satellite periods.

I have added a citation to Gmez-Escolar et al. as suggested. While it is possible that this reflects some true shift of the statistical seasonality of sudden warmings, it is also completely consistent with the null hypothesis that this is a result of sampling variability from a stratospheric climate that has not changed. This can be regarded as a source of decadal variabiltiy, but the statistics being what they are, it seems most reasonable to stick with this null hypothesis. This is also relevant to a point raised by Reviewer 3; I have added a bit of text in the discussion on this point as well.

Page 7 lines 23-25: Maybe I am getting something wrong but the largest spreads, at least for the zonal wind, are found in the Northern Hemisphere.

This is a good point - this is true of the upper stratosphere winds. I have been more careful to mention this in the text.

Page 8 lines 11-12: Please indicate why you are selecting different levels for the stratospheric field in the Northern and Southern Hemisphere.

This is reasonable; I had chosen a lower height to see if there was a level where the Southern Hemisphere might be better constrained, but having looked at this again it does not make a big difference and so I have changed the figure to show 30 hPa for both hemispheres to avoid having to explain any differences.

Page 9 lines 20-25: I think the author should be careful with the description of the results in this paragraph. For instance, some fields that are indicated as not shown (for T in JJA, for DJF u) are in fact shown and some others described as shown are not (for T in DJF, for u in JJA). It would also help if a reference to the plots is included in each case too.

This has been corrected and some additional explicit referencing of figure captions has been added.

Page 10 lines 32-35: I might agree that data of 1950s may be of interest, but the results for NCEP/NCAR reanalysis for that decade are not shown in Figure 9.

This paragraph has been removed.

Technical comments

Page 4 Line 21: then than

Page 4 Line 31: I think it would be better to write from 1958 to 2016.

Page 5, equation 2: in the second sum the upper limit should be N r instead of N s .

Page 5 equation 2: Please define N t

Page 7 line 15: I think it is the winter upper stratosphere.

Page 7 line 25: in many regions in in many regions it

Page 8 line 25: Southern Hemisphere

Page 10 line 26: Please delete at.

Page 10 line 30: Please include) after 9.

Page 11 line 6: reduced reduce.

These have all been corrected.

Response to Reviewer 3

This is an interesting and well thoughout study about the potential value of earlier, pre-satellite era reanalysis records. It is important to quantify the potential value of this earlier period, as it is a major undertaking for a reanalysis center to provide pre- satellite reanalyses. With the exception of JRA-55 (and the ERA5 analysis, currently in production), most of the state-of-the-art full input reanalyses do not begin until 1979 (ERA-I, MERRA, CFSR) or 1980 (MERRA2).

I recommend publication of the manuscript pending consideration of the comments below. They are mostly minor, in that I leave them to the authors discretion, but I hope that responding to them would improve the impact of the paper. (An exception is that the author does need to better define a few things, to ensure the results are reproduceable. But this will be easy to do.) My more philosophical question about the proposed metric for assessing the value of earlier reanalyses (see below) is perhaps trickier to fully answer, and might be something for future work. I think that the contributions of this paper are already worthy of publication. Given that it could be a subject for future research, Ill sign this review, as I would welcome discussion with author.

Edwin Gerber

Thank you for your comments.

General comments

1) A few key elements of the procedure were not sufficiently documented. In particular, how were the SSW dates set, and how were the events classified in the spilits or displacements. I suspect this was done within the S-RIP Chapter 6 framework, assembled by Amy Butler. If so, I am not sure how to properly cite this information at this time, though they will be published. In any case, to reproduce these results, the reader does need to know the dates, and some insight on how they were obtained.

This is the case; I have added some details to the text regarding the event definitions, but have not added an explicit list of dates. This could easily be done if deemed necessary.

2) It would help the reader to adopt a consistent use of the nomenclature ”full-input”, ”conventional-input”, and ”surface-input” throughout the paper. I appreciate that terms evolved in parallel to this research, but as a result of this time mismatch, they appear inconsistently through the text.

These terms have now been explicitly defined (in reference to Fujiwara et al. 2017) and used more consistently throughout the text.

3) I very much appreciate the central result of the manuscript: equation (3) and surrounding discussion, which seeks to quantify the value of earlier records. I was admittedly surprised, however, that the metric indicates that there is considerable value to much of the data in the austral hemisphere, where we know that the large scale circulation is not consistently captured by the reanalyses. (In Gerber and Martineau, 2018, for example, we found that the southern annular mode indices in JRA-55, ERA-40, and NCEP-R1 share only a small fraction of the variance during the pre-satellite period, indicating that there is very little consensus on the large scale state of the austral hemisphere on synoptic time scales.)

I think the key is the assumption that reanalyses properly capture the dynamical uncertainty, σ_d, in both the satellite and pre-satellite periods. I think this effectively implies that we trust their climatological values and variance, even if the become untethered to observations.

To make my concern clear, consider the extreme case where the reanalyses are perfect in the satellite era ($\alpha_s \to 0$) and know absolutely nothing about the state of the atmosphere in the radiosonde era ($\alpha_r \to \sqrt{2}$). In this case, $f \to 2$ and $\delta \to (1 - 2\beta) / (1 + (1 - \beta)2)$

When β becomes small (¡0.5), you would still conclude that there is value in the reanalysis, even though it knows nothing about the state of the atmosphere. (The "real" β is about 0.6, so in this limiting case delta would be negative, and you would concluded there is no value in earlier records). But given that α_s is not zero, and there is some limited skill in the radiosonde era, its not hard to see why delta is positive. And by this logic, there would be considerable value in using the entire record from ERA-20C, where beta drops below 0.5!

My intuition if we want an observationaly constrained estimate of the uncertainty, then we should only include the information from the earlier period when $\alpha_r < 1$. That is, when uncertainty in the renanalyses reaches the level of dynamical uncertainty, then we can argue the reanalyses are sufficiently untethered from the real atmosphere to provide any additional information than you could obtain from simply running a forecast model untethered to observations.

I havent thought this through enough to provide a good way to quantify the value of events when $\alpha_r < 1$. It helps me to think of this interms of events (as with the SSW composites shown by the author.) Suppose you have N events from the satellite record. Looking at past events, the idea would be to quantify the additional information content of each radiosonde period event on an event-by-event basis. When the spread between reanalyses for an earlier event is equivalent to the spread between events in the satellite period ($\alpha_r = \alpha_s$), the event is clearly of complete value ($\delta = 1$); it should be added fully. Now your composite is based on N+1 events, and the uncertainty drops accordingly.

If the spread between the reanalyses for the event, however, becomes equivalent to the climatological/dynamical spread ($\alpha_r = 1$) then I feel that theres no additional information to be gained than if you simply ran a free running model: this event should be given zero value. I am just not sure how to develop a meaningful way to interpolate inbetween these cases.

Perhaps the central issue here is that (3) measures the contribution towards reducing the variance of the sample mean from two samples drawn from a population with the same mean. In essence it is staring from the assumption that the forecast model has the same climatology as the real atmosphere, and if this really was the case then it would be worth including data from a period when the model was completely untethered from observations, since the longer time period would still act to reduce sampling uncertainty. So this metric does not tell us everything we need to know about how much information comes from the real atmosphere versus how much from the forecast model.

The sensible further criterion is whether the forecast model is actually following fluctuations that occured in the real world. In this case the spread between multiple reanalyses is being used to estimate α (in either case), and in the case where they all have the same variance as the real atmosphere but are fully independent realizations, σ_o should approach $\sqrt{2}$ times the variance of the reanalyses. (This is laid out a bit more explicitly in the text now.)

So long as σ_d in the reanalyses isn't too far off, $\alpha \to \sqrt{2}$ (in either era) is what one expects if the forecast model is just doing it's own thing. It's probably reasonable to set the bar rather lower than $\sqrt{2}$, but how far is a matter of judgement or of further criteria. The colouring adopted in Fig. 6 used 0.1, 0.3, and 1.0 as thresholds and so reflects the value of 1 you've suggested, but this is to some extent arbitrary.

As discussed further below, while this discussion emphasizes the difficulty of where to draw the line between including or not including the radiosonde era, a more important take away here is that so long as the α's are small relative to σ_d (as is particularly the case for zonal wind in the NH stratosphere), it doesn't matter how much bigger α_r is relative to α_s, it's still worth including the radiosonde period.

Another criteria one could think of is the possibility of systematic bias, but this kind of reanalysis intercomparison can't speak directly to that. The text gives a rough argument for when this can be neglected relative to σ_d. (The discussion of Fig. 10 is also relevant.)

From an event point of view, if the reanalysis is actually capturing the same event as the observations, δ is the relevant measure and serves to quantify your example.

In view of these points, I've significiantly reworked the discussion of (3) to clarify these considerations, and I've added more emphasis to the importance of $\alpha_{r,s}$ being small.

4) I appreciate that the comment above is weak on specific suggestions. To be more concrete, I would have appreciated more discussion of the different limits around equation (3). The limit where α_s is small and α_r approaches sqrt(2) was interesting to me, as it drove home this issue of whether we ought to trust a good model that is untethered to reality.

Another problematic limit is $\alpha_r = \alpha_s$. Here, you always use more data, even if its all untethered to reality. (Based on my arguements above, the value of the reanalyses should be zero when α_r or α_s approaches 1.)

The problem here is really just the same as above; when either α_r or α_s approaches $\sqrt{2}$ this is indicative of the reanalyses becoming untethered to the observations in the respective period. If they are both roughly equal, the two eras contribute equally to reducing the uncertainty as should be the case.

And not to be overly critical, Figure 4 was not easy to interpret. Consider using color or more simply marking the contours. (I know that I should have realized that diagonal is 1 by definition, but it took me time at first reading.) I also think that its inappopriate to show such a range. Once α_s or α_r reach $\sqrt{2}$, nothing is tethered to observations, and I dont see how δ is meaningful for values beyond this point.

This is a good point, the range shown has been reduced accordingly. I have also added labels to the contours.

5) There is a paper that can be cited for the Martineau data set: 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, 19251941, doi:10.5194/essd-10-1925-2018, 2018b.

This has been added.

Also, its my understanding that MERRA2 has a DOI that should be cited, as its very important for them to justify resources. I find the situation problematic, in that you got the data from a different source (which did cite this doi), but perhaps you could add the doi to the data section.

As you say, given that the dataset I've used is not the reanalysis center itself, it seems more appropriate to cite the Martineau paper/dataset, especially since there is a citation for each reanalysis already.

Small comments 2:3-8 This would be a good time to differentiate and define full-, conventional- and surface-input reanalyses.

I've done so a bit later, when discussing the reanalysis datasets in Section 2.

2:21

citep[e.g.,][] (Also, I think that perhaps one should include the comma on e.g., since if you spelled out the phrase, it would be: for example, Matsuno 1971. But perhaps this is a case where American English is different from British.)

Done.

2:34 I might break this off as a full sentence, instead of using the semicolon.

Done.

3:11 I appreciate why the author states that they are constrained *primarily* by surface observations (as the reanalyses are given changes in radiative gases, etc.), but this sentence seemed a bit to vague.

I have reworded this to emphasize that upper-air observations are not assimilated by these products.

3:25 consider rephrasing this sentence: I understood it completely, but had to re-read it a few times

I have reworded, hopefully it is clearer.

4:16 This would be a place to explain how the dates were set, and how splits and displacements were classified, or at least point the reader to the necessary information.

This has been done.

4:24-26 This is a fascinating/perplexing result. I think it makes sense, though: ERA- 20C does a good job of getting SSWs, but since it only gets the dates right for half of them, you are better off treating it as a free running model (i.e. not fixing dates to reality) than trying to make it conform with what actually happened in our atmosphere. ERA-20C provides the challenge to your metric in equation 3: I do think you would argue that its worth while using the entire record, even if just assume it knows nothing about the actual state of the atmosphere. Thats what motivated my thoughs on comment 3 above.

If I've understood you correctly, this is a case where (3) would not see a difference between the satellite era and the radiosonde era, and would therefore weight the whole record evenly. As you suggest, this is a case where the second criterion is important; for the cases shown in Figs. 6, 8, and 9, α is generally smaller than expected if ERA-20C was completely untethered to observations, but still or the order 1.

4:29 splits and displacements need to be defined

The method of Lehtonen and Karpechko has been used; this is now stated explicitly. The reduction of CI's is the same for other definitions that I've tried.

5:1 second half of the line is awkwardly phrased

Reworded.

5:17 Im concerned that the zonal mean wind at 60 N and 10 hPa is decidely not Gaussian, and rather skewed towards negative values.

This is true, and so for climatological averages this could be an issue - but as is stated in the next sentence, the central limit theorem helps out. It's also less clear this is as important for things like composites which may give rise to more gaussian statistics. More generally, getting into a detailed discussion of the corrections one might have to adopt in the presence of non-gaussianity seems a distraction at this point.

5:18. There is a sentence between when you introduce σ_d and σ_o and define them. Consider moving the first sentence of the next paragraph up, to define the variables, before discussing the central limit theorem.

This would move the comment about non-gaussianity even further from the example given for X; but I've slightly changed the discussion to bring the definition closer to the introduction of these terms.

6:10 In the limit where the reanalysis error is small relative to the dynamical uncertainty, isnt f small, and delta about equal to 1?

To $O(\alpha_{r,s}^0)$, yes. To leading order in $\alpha_{r,s}$ the expression is as given. β enters at the next order of the expansion (as $1-\beta$, in fact), and this term is less than a 10% correction to δ for δ as small as 0.5, so long as beta isn't too small. See also the response to a later comment.

8:20-24 This sentence is long. Consider breaking at the ;, and then being more clear what agreement you mean to refer to.

Reworked as suggested.

Fig. 8: I assume the Fourier analysis was done on the deaseasonalized winds, as theres no discernable annual cycle peak here!

Yes, this is the case. This is now stated in the text.

10:21 and Figure 7 e,f I am confused how you can estimate δ without knowing β. It only seems to decouple from β when f is small. And in this limit, δ would be close to 1 (and its sort of a trivial result: you trust everything.) In the figure, the value of delta varies considerably (changing sign!) so you must have some finite value of β. What is 0.6?

The calculation was done assuming a value of 0.6 for β; in light of the discussion above around the importance of α itself, I've changed this figure to show α estimated on a month-by-month basis. This also removes the question of *the dependence on β.*

11:1-2 This sentence could be split up, giving you two sentences, enough to justify a paragraph!

Done.

11:13-15 It was hard for me to see this important result. At some latitudes (c. 65 in panel a of Fig 10, or near 75 in panel b), the uncertainty bounds on the full record were larger than than for the satelite record. So clearly there wasnt always a 20% reduction. Im was also rather struck by the fact that the dashed curved in panel b of Figure 10 approaches the edge of the confidence interval on the "all" composite. Does this mean that they were almost statistically different, or would this only apply when the confidence intervals themselves separate.

The confidence intervals are themselves statistical estimates which will only approach the 20% reduction in a probabilistic sense; so yes, while there are regions where this does not hold, it is broadly the case that there is a reduction of this order. I have made it clearer that this reduction is not universal. As for the differences, perhaps you meant the blue dashed curve in panel (c)? In any case both are in fact within the confidence intervals given. Although I have not carried out an explicit test of statistical differences this seems unlikely.

From a practical standpoint, if I wanted to ask whether my model was signficantly different from our best estimate of observations, which error bound should I use?

On the weight of evidence presented (and reviewed) here, the bounds based on all data available. This has been emphasized in the text.

11:30 Might be good to emphasize "has been quantified in equation (3)."

Done.

12:2 Related to some comments above, when the dynamical uncertainty dominates, doesnt this imply that you trust everything?

Yes - as discussed above this is a key message. (Although a better way to think of it might be that everything is valuable to reducing this uncertainty). I have re-emphasized it in the text.

12:15 An opportunity to use surface-input nomenclature.

Taken.

15:18 I am not sure how to see this in Figure 9. Doesnt the fact that δ is consis- tently greater than 0 for ERA-20C imply that theres always value to be found from this reanalysis?

I have rewritten this paragraph emphasizing points that have already come up in the discussion above.

15:26-30 Could be opportunity to highlight that your message has been heard, and ERA5 hopes to go back to 1950.

Done, also following suggestion of reviewer 1.

Fig 4: See my general comment (4) above. I think this figure could be improved.

I have reduced the range shown and added labels to the contours.

Fig 5: Your notation differs a bit here, σ_d vs. σ_{dyn}. Its clear enough for the reader, but consistency is best.

Fixed.

Fig 7: caption has the wrong symbol. I would have appreciated more detail here in how the bottom panels were computed.

The caption has been corrected.

Fig. 8: I find that the log scale makes comparison very difficult. Would it be possible to show the ratio of the differences in the power spectra? This is a number that would presumable vary from 0 to about 2 for all timescales. (It would be 2 if the limit that the models become untethered from observations. I guess it could become larger if there are systematic biases.)

I've taken the suggestion to show ratios (and added a line at 2 for reference). I do think this makes the figure and its interpretation clearer.

If nothing else, the reference time series of JRA-55 gets buried by the other lines: consider bringing it up to the top. (If you produced this plot with matlab, but want to keep it first in the legend, a solution is to just print it again.)

Following your first suggestion, JRA-55 is no longer shown.

Figure 10: Are these 95% confidence intervals?
Yes, this has now been clarified in the text.

On the value of reanalyses prior to 1979 for dynamical studies of stratosphere-troposphere coupling

Peter Hitchcock[1]

[1]Laboratoire de Météorologie Dynamique, École polytechnique, Palaiseau, France

Correspondence: Peter Hitchcock (peter.hitchcock@lmd.polytechnique.edu)

Abstract. Studies of stratosphere-troposphere coupling, particularly those seeking to understand the dynamical processes underlying the coupling following extreme events such as major stratospheric warmings, suffer significantly from the relatively small number of such events in the 'satellite' era (1979 to present). This limited sampling of a highly variable dynamical system means that composite averages tend to have large uncertainties. Including years during which radiosonde observations of the stratosphere were of sufficiently high quality ~~could substantially extend~~ substantially extends this record, ~~potentially~~ reducing this sampling uncertainty by up to 20%. Moreover, many open questions in this field involve aspects of tropospheric dynamics likely to be better constrained by 'conventional' (i.e. radiosonde and surface-based) observations.

Based on an inter-comparison of reanalyses, a quantitative case is made that for many purposes the improved sampling obtained by including this period outweighs the reduced precision of the reanalyses in the Northern Hemisphere. Studies of stratosphere-troposphere coupling should therefore consider the use of this period when using reanalysis data~~, and the community should advocate for continued attention to be focused~~. These results also support continued attention on this period from centres producing reanalyses.

1 Introduction

One of the central challenges to the detailed study of the large-scale coupling between the stratosphere and the troposphere is the relatively limited record of high quality, global observations. In the absence of more insightful modes of analysis, quantifying the dynamical processes relevant for the coupling requires large samples to isolate them from unrelated dynamical variability. Despite the availability of nearly four decades of global satellite-based observations, the length of the observational record remains a fundamental limitation to this statistical approach. This is demonstrated explicitly here, as well as by another closely related contribution ~~(?)~~ (Gerber and Martineau, 2018) to the SPARC Reanalysis Intercomparison Program (SRIP; Fujiwara et al., 2017).

The coupling between the stratosphere and the troposphere remains a significant source of uncertainty in projected climate changes over the coming century (Manzini et al., 2014; Simpson et al., 2018), as well as an important source of skill in seasonal forecasting (Sigmond et al., 2013). Global models exhibit a diversity of stratospheric circulation (Manzini et al., 2014) and variability (Charlton-Perez et al., 2013; Taguchi, 2017), and of tropospheric responses to stratospheric variability (Hitchcock and Simpson, 2014). Observations of the true circulation can be used to identify which models are correctly representing these processes, but this relies on comparing the time-averaged behaviour of the models to the observations, and the large interannual variability in the observed circulation means that the sampling uncertainty remains large. Accounting for sampling error in such large-scale dynamical phenomena is a major concern for many other dynamical questions, including identifying regional signals of climate change and teleconnection patterns ~~(e.g. Deser et al., 2017)~~(e.g., Deser et al., 2017).

Studies of observed stratosphere troposphere coupling often rely on reanalysis products, which combine a wide range of observations with global forecast models (see Fujiwara et al., 2017, for a comprehensive discussion). Two of the older products, ERA-40 and NCEP-NCAR R1, begin in 1957 and 1948, respectively, dates which coincide with significant extensions of the global radiosonde observing network. Many more recent products (ERA-Interim, MERRA, MERRA-2, CFSR) by contrast cover only the period from 1979 onwards, after the availablility of sounding data from Microwave Sounding Unit (MSU) and Stratospheric Sounding Unit (SSU) instruments. It is convenient to label the period after 1979 the 'satellite' era, ~~that is, the period after 1979.~~ though it is worth noting that a number of satellite data products exist prior to 1979, as discussed by Uppala et al. (2005). Amongst the more modern products only JRA-55 begins prior to the satellite era, in 1958. However, both ERA-5 and JRA-3Q, two newer products unavailable at the time of writing, are expected to cover the pre-satellite era as well.

For the purposes of the present work, the 'radiosonde' era will ~~be used to~~ refer to the period from 1958 to ~~1978.~~ 1978, although radiosonde data exists prior to this period and continues to be important afterwards. There is no general consensus amongst studies of stratosphere-troposphere coupling as to whether to include the radiosonde era. This is complicated by the fact that the coverage of ERA-40 ends in 2002, leaving out the most recent (and best-observed) decade and a half. Some studies have made use of the older reanalysis products ERA-40 and NCEP-NCAR R1 alone (Charlton and Polvani, 2007; Mitchell et al., 2013) while others consider exclusively the satellite record (Dunn-Sigouin and Shaw, 2014; Kodera et al., 2015; Birner and Albers, 2017). Still others choose to merge multiple reanalyses, using an older product for the radiosonde era and a more modern product for the satellite era (Hitchcock et al., 2013; Lehtonen and Karpechko, 2016). The value of JRA-55 as a single modern product that spans both the radiosonde and satellite eras is thus evident, (and as such it will be privileged in the analysis that follows) but the question remains whether the observational record during the radiosonde era is of 'sufficiently' high quality to be worth considering.

~~Given that the~~ The first identification of a stratospheric sudden warming is credited to Scherhag (1952) and ~~that~~ much was known about their dynamics prior to the availability of a long satellite-based observational record ~~(Matsuno, 1971; Labitzke, 1977; McIntyre~~ (e.g., Matsuno, 1971; Labitzke, 1977; McIntyre, 1982), largely on the basis of radiosonde observations~~,~~. Moreover, a successful five-day forecast of the sudden warming that occurred in January 1958 initialized from ERA-40 has been demonstrated (Simmons et al., 2005). All of this suggests that the observational record prior to 1979 ~~would seem to be of clear value~~ is of real value in constraining the behaviour of the coupled stratosphere-troposphere system around stratospheric sudden warmings.

The immediate goal of this work is to evaluate the representation of a number of quantities of interest to the problem of stratosphere-troposphere coupling in the radiosonde era, in view of coming to a more quantitative assessment of their value. For the Northern Hemisphere the arguments given below clearly indicate their value. However, since this judgement depends on the specific quantity of interest, a broader goal is to discuss how to answer this question more generally. Indeed, the same arguments should apply to the study of many other features of the large-scale atmospheric circulation, particularly of those phenomena with large spatial scales and characteristic timescales of the order of weeks to months. The same approach could also be applied in principle to the period prior to 1958, although no effort has been made to do so here.

This evaluation is based on the availability of multiple reanalysis products. Since in general the different reanalyses assimilate subsets of the same observational record into distinct forecast models, the level of agreement provides a simple measure of how strongly the observations constrain the quantity in question. This method has caveats in that the underlying forecast models may share biases that result in them getting consistently wrong answers; more . More critically, the availability of only one modern reanalysis product that covers the radiosonde era (and assimilates radiosonde data) means that this comparison must be based in part on older reanalyses with known deficiencies (e.g. Long et al., 2017)(e.g., Long et al., 2017). Nonetheless, as will be argued below, the agreement is close enough in the Northern Hemisphere to suggest that this period has real value for carrying out many classes of dynamical studies. This is broadly consistent with the conclusions of Gerber and Martineau (2018) and of Hersbach et al. (2017), which explicitly examined the value of upper-air observations over the period 1939 to 1967 in an experimental reanalysis product.

The outline of this paper is as follows. The reanalysis data considered here is described in Section 2. Section 3 presents, as an initial example, a discussion of the time series of zonal mean zonal wind at 10 hPa and 60° N that is central to the identification of major sudden stratospheric warmings. Section 4 presents more general criteria for determining when the radiosonde era should be included. These criteria are then discussed in Section 5 as they apply to wider variety of zonal mean quantities, including fluxes of heat and momentum that are relevant to stratosphere-troposphere coupling. Section 6 presents conclusions and a discussion.

2 Reanalysis data

Zonally averaged output from the 12 reanalysis products listed in Table 1 are considered here. Of these reanalyses, five (JRA-55, NCEP-NCAR, ERA-40, 20CR v2, and ERA-20C) include the period from 1958 through 1978. Two reanalysis products (20CR v2 and ERA-20C) extend further back but are constrained primarily by surface observations. The do not assimilate upper-air observations; following the nomenclature of Fujiwara et al. (2017), these will be referred to as ‘surface-input’ reanalyses, in contrast to ’full-input’ reanalyses. A third category is ‘conventional-input’, the sole present example being the JRA-55C productis also . This is noteworthy in this context as it assimilates only ‘conventional’, that is to say, non-satellite based, observations. It therefore provides a means of estimating of the additional value of incorporating the satellite observations. A useful comparative description of these reanalysis products including details of the underlying forecast models, the observational datasets assimilated, and the assimilation techniques used can be found in Fujiwara et al. (2017). The data used here has been re-gridded to a uniform latitude-pressure grid, and is available for download (Martineau, 2017)described in Martineau et al. (2018).

Anomalies are computed from climatologies based on the years 1981 through 2001. These years are chosen since they are included in all of the reanalysis products under present consideration. Leap years are handled by omitting July 1st so that all years are treated as 365 days long. These climatologies (computed for each reanalysis) are used regardless of the period under consideration.

3 Sudden Stratospheric Warmings

As an initial example, Fig. 1a shows time series of zonal mean zonal wind at 60°, 10 hPa from the JRA-55 reanalysis for a set of 36 stratospheric sudden warming events, identified following Charlton and Polvani (2007). The central dates (lag 0) of the events are defined by when the wind at this grid point reverses from westerly to easterly, so all of the time series pass through ~~0~~ zero at this point. However, the inter-event variance of the winds ~~grows rapidly both~~ is large both immediately prior to and shortly after the central date. This spread is only to a weak degree the result of the timing of the event within the cold season; a similar plot of anomalies from the climatological mean shows very similar growth in the inter-event spread (not shown). As a result of this large dynamical variability, the composite mean has a large sampling variability independent of the quality of the observations or the forecast models underlying the reanalysis products.

In contrast, Fig. 1b shows the same time series from all twelve reanalysis products for a single event that occurred on 21 Feb 1989. The inter-reanalysis spread is in general much smaller than the inter-event variability emphasized in Fig. 1a. An exception to this is the surface-input reanalyses, ERA-20C and 20CR v2~~which assimilate only surface observations~~. JRA-55C, which does not assimilate satellite observations, is notably indistinguishable from other reanalysis products, suggesting that satellite observations are not required to closely constrain these winds.

Although there are far fewer reanalysis products that include the radiosonde period, Fig. 1c shows that the three reanalyses spanning this period which assimilate radiosonde observations (JRA-55, NCEP-NCAR, and ERA-40) exhibit a similarly close agreement, showing only a somewhat larger spread across reanalyses than in the satellite period. This again suggests that the radiosondes are providing a strong constraint on the flow, and that as a result the events that occurred during the radiosonde era are of significant potential value for constraining our knowledge of the composite mean evolution of sudden warmings.

Since sudden stratospheric warmings are typically identified by the date on which this wind reverses sign, these slight differences in reanalyzed winds can lead to the identification of central dates which differ by a day or two, and in some cases can lead to an event being identified in one reanalysis but not in others. This sensitivity is a generic feature of thresholds in the event definition, not of the particular choice of definition.

This leads to difficulties with comparing composites of events in different reanalyses: because of the large inter-event variability, the exclusion of even just one event from a given reanalysis composite mean can produce differences in the composite mean that easily overwhelm the differences in the reanalyzed flow itself. Thus small differences in the identification of events can 'alias' into relatively large apparent differences in the overall composite evolution.

Similar considerations preclude the direct comparison of composite averages of satellite-era and radiosonde-era events: they differ, but not evidently by any more than should be expected due to this dynamical sampling uncertainty. To isolate the intrinsic differences between reanalyses from this aliasing of sampling variability one must instead consider a fixed set of events across all reanalyses. This ~~approach is followed here~~ is done here by selecting the date where the event fell in the majority of the available reanalyses, following the S-RIP Chapter 6 analysis of stratosphere-troposphere coupling.

These points are illustrated in Fig. 2, which demonstrates that composites of events across reanalyses agree better when a fixed set of dates is taken ~~then~~ than when event dates are chosen individually for each reanalysis. This is true of the full-input analyses for both the satellite era and the radiosonde era.

In contrast, the ~~surface~~ surface-input reanalyses (ERA20c and 20CR v2) generally agree better with the composites when event dates are chosen per-reanalysis~~–~~, particularly around the central date of the event. This suggests that while the surface observations are sufficient to constrain the stratospheric flow to some extent, the break down of the stratospheric vortex is ~~still to a significant extent~~ also significantly determined by the behaviour of the forecast model in these products.

Considering a list of fixed event dates provides a useful starting point for quantifying the additional information contained in the radiosonde era. Using the fixed set of event dates as a basis, Fig. 3a shows estimates of the overall frequency of stratospheric sudden warmings for the satellite era alone and for the full 1958-2016 era, as well as for split and displacement events. The month-by-month frequency is shown in Fig. 3b. Confidence intervals in all cases are estimated with a bootstrapping procedure: N years are selected from the period from ~~1958-2016~~ 1958 to 2016 with replacement, and the events that occurred in these N years are then used to compute event frequencies, counted multiple times for those years that are selected more than once. For the satellite era $N = N_s = 32$, while for the total period $N = N_t = N_s + N_r = 53$. This whole processes is repeated 10000 times, and the bounds of the confidence intervals are taken to be the 2.5th and 97.5th percentiles.

As expected from the central limit theorem, the confidence intervals are ~~sealed by~~ reduced by a factor very close to $\sqrt{N_s/N_t}$~~, amounting~~. This amounts to about a 20% reduction~~. This improves the~~, providing a stronger observational constraint on the climatological frequency of sudden stratospheric warmings. A similar reduction is obtained for the occurrence frequency of splits and displacements, classified following Lehtonen and Karpechko (2016), as well as for the seasonal distribution of events.

Since the bootstrapping is based on the entire record, the confidence intervals for the satellite era are not centered on the mean frequencies. The use of the longer baseline results in a slight shift of the seasonal peak, suggesting that in the long term, January events are in fact more frequent than February events, in contrast to the February peak obtained using the satellite period alone. This difference in apparent seasonality has also been discussed by Gómez-Escola et al. (2012). These changes could in principal be a result of some longer term trend or decadal variability external to the stratosphere, but they are fully consistent with the null hypothesis of sampling variability from an unchanged underlying seasonality. In this latter interpretation, the full record therefore represents a modest but useful strengthening of the observational constraints on these statistics.

4 ~~A~~ Statistical ~~Criterion~~Considerations

Despite these promising examples, one should expect in general that the quality of the reanalyses are not as high during the radiosonde era as during the satellite era. In this light one might regard the ~~improvement~~ reduction of 20% in the confidence intervals found in Fig. 3 to be an upper bound~~on the degree of improvement~~. While errors in the reanalyses will in general arise from both observational uncertainty as well as from ~~errors in~~ uncertainty arising from the underlying forecast model and assimilation process, these will be considered together here as ‘reanalysis’ uncertainty.

A simple ~~means of quantifying this improvement~~ way to quantify the potential improvement from including the radiosonde era is to treat the reanalysis and sampling uncertainty as uncorrelated, gaussian variance, and consider the effect on the sample mean of ~~an inhomogeneous set~~drawing from two periods with different variances. More explicitly, we consider some physical observable X (for instance, the zonal mean zonal wind at 10 hPa and 60° N) to be modeled by a normally distributed random variable with mean μ and variance ~~$\sigma_d^2 + \sigma_o^2$~~σ^2. Since we are interested in the statistics of the sample mean, the central limit theorem ~~would in principal allow~~ in principal allows the assumption of ~~normality~~ gaussianity to be relaxed, but the role of non-gaussian statistics will not be explicitly considered.

~~The~~ We further assume that the variance consists of two uncorrelated components ~~: one~~ $\sigma^2 = \sigma_d^2 + \sigma_o^2$: the first, σ_d^2, arising from the dynamical variability of the atmosphere, ~~σ_d^2, and the other~~ and the second, σ_o^2, from the reanalysis uncertainty~~, σ_o^2~~. We further consider two sets of observations of this variable, one of N_s samples with smaller reanalysis error representing the satellite era, with $\sigma_o = \sigma_s$, and one with N_r samples and relatively larger reanalysis error representing the radiosonde era, with $\sigma_o = \sigma_r$. We take the dynamical variability to be constant across both samples. The variance of a sum of independent random variables is the sum of the variance of each variable; hence the variance of the sample mean during the satellite era is

$$\mathrm{Var}\left(\frac{1}{N_s}\sum_{i=1}^{N_s} X_i^s\right) = \cancel{\frac{1}{N_s}\sigma_d^2 + \sigma_s^2}\,\frac{(\sigma_d^2 + \sigma_s^2)}{N_s}, \tag{1}$$

while that of the sample mean over the entire period is

$$\mathrm{Var}\left(\cancel{\frac{1}{N_t}}\,\frac{1}{N_s + N_r}\left(\sum_{i=1}^{N_s} X_i^s + \sum_{i=1}^{\cancel{N_s}\,N_r} X_i^r\right)\right) = \cancel{\frac{1}{N_t^2} N_s \sigma_d^2 + \sigma_s^2 + N_r \sigma_d^2 + \sigma_r^2}\,\frac{N_s\left(\sigma_d^2 + \sigma_s^2\right) + N_r\left(\sigma_d^2 + \sigma_r^2\right)}{(N_s + N_r)^2}. \tag{2}$$

Here the superscript on X indicates the ‘era’ from which the sample is drawn (and thus its variance).

A ~~simple~~ first criterion for including the both periods is that the standard deviation of the sample mean should be reduced relative to that obtained from the satellite era alone. As argued in the previous section, if the reanalysis error of the two periods are equal ($\sigma_r = \sigma_s$), the standard deviation of the mean when the whole record is considered will be reduced by a factor $\sqrt{N_s/(N_s + N_r)}$. If the reanalysis error of the two periods differ, some straightforward manipulations of the formulas above can be used to show that the factor can be written $\sqrt{N_s/(N_s + \delta N_r)}$, with

$$\delta = \frac{1 - \beta f}{1 + (1 - \beta) f}, \qquad f = \frac{\alpha_r^2 - \alpha_s^2}{1 + \alpha_s^2}. \tag{3}$$

Here $\alpha_{s,r} = \sigma_{s,r}/\sigma_d$ is the ratio of the reanalysis standard deviation in each respective period to the dynamical standard deviation, and $\beta = N_s/N_t$ is the length of the satellite era as a fraction of the total length of the record. For the observational period considered here, $\beta \approx 0.6$.

The factor δ can be loosely interpreted as an efficiency factor for the sampling during the radiosonde period. Since it depends on the number of observations in both periods its value will in general change (through β) with the size of the sample; however, in the limit that the reanalysis error in both eras is small compared to the dynamical error, $\delta \approx 1 - f = 1 + \alpha_s^2 - \alpha_r^2$, in which case its value is independent of the sample size. This result, central to the argument of this work, indicates that even if the reanalysis uncertainty in the radiosonde era is much larger than the reanalysis uncertainty in the satellite era, δ will be close to 1 so long as the dynamical uncertainty dominates both.

Figure 4 shows values of δ as a function of α_r and α_s for three values of β. One can note several properties of this factor. Firstly, so long as the reanalysis uncertainty in the radiosonde period is larger than that in the satellite era ($\alpha_r > \alpha_s$) δ will be less than 1, with $\delta = 1$ if and only if $\alpha_r = \alpha_s$. Secondly, δ can be negative for sufficiently large values of α_r, although this threshold depends on the value of β. For the present observational record (Fig. 4b), when α_s is small this occurs only when α_r is somewhat larger than 1, that is, when the reanalysis uncertainty is somewhat larger than the dynamical uncertainty. This threshold occurs at smaller values of α_r when the satellite era comprises a larger fraction of the record, as can be seen from comparing the three panels. as β decreases, so that, for marginal cases, the value of the radiosonde era in reducing overall uncertainty will decrease with time as a longer record of higher quality observations becomes available.

In practice, Secondly, δ remains close to 1 if $\alpha_r \approx \alpha_s$. Because this statistical model assumes that both periods are drawn from populations with the same underlying mean, it assigns equal value to both periods, regardless of how large the reanalysis uncertainty σ_o is estimated here from the statistics of differences between different reanalysis products, while is relative to the dynamical uncertainty. In practice, the dynamical variability is estimated σ_d is estimated here from the interannual variability of the field in question. As discussed above, the reanalysis uncertainty thus includes both observational uncertainty as well as errors in the forecast model and the assimilation processThe reanalysis uncertainty σ_o is estimated from the statistics of differences between different reanalysis products: more precisely as the time mean of the standard deviation across reanalyses. If the observations are not constraining the flow in a significant way, the reanalysis product will reflect the dynamics of the underlying forecast model and the flow across the various reanalyses will become uncorrelated. If In this case, assuming that the forecast models produce reasonably accurate dynamical variability, the estimate of σ_o should approach $\sqrt{2}\sigma_d$, that is, $\alpha \approx \sqrt{2}$. To see this, consider the time series of an observable from a given reanalysis X_i as the sum of the true atmospheric evolution X_a and a correction x_i. If the standard deviation of the differences should approach $\sqrt{2}$ times the dynamical variability; that is , $\alpha_{r,s} \approx \sqrt{2}$. forecast model is correct, X_i has the same standard deviation as X_a. When these two components become decorrelated, the correction x_i will be the difference between two uncorrelated timeseries with standard deviation σ_d. Since X_a is independent of the reanalysis, the standard deviation across reanalyses will therefore be $\sqrt{2}\sigma_d$.

This suggests a second criterion; if the variance of these differences approaches this value, it suggests that : if α_r (or α_s) approaches $\sqrt{2}$ the observations are not providing any significant constraint on the fluctuations, and thus that the variability in the reanalysis is arising purely from . In this case we should not regard the reanalysis as providing any kind of estimate of the true behaviour of the climate system and this part of the time series should not be included. To avoid influence of the forecast modeldynamies, one might reasonably require α to be significantly less than $\sqrt{2}$.

An important assumption that has been made is that the reanalysis uncertainty is dominated by a stochastic component that is uncorrelated across the samplesin time. One can imagine easily suppose the presence of systematic errors that remain relatively fixed in time, differing only when the assimilated observations change in a substantial way. Such a systematic error will not be reduced by a larger sample size; if such an error ϵ is present during the radiosonde era, its contribution to the overall uncertainty will be $\epsilon(1-\beta)$. However in the case that the dynamical sampling error dominates the random component of the uncertainty, ~~the~~ this systematic error can still be ~~negligible if $\epsilon < \sigma_d/\sqrt{N_t}$~~ neglected if $\epsilon \ll \sigma_d/\sqrt{N_t}$.

Since the dynamical standard deviation is in general a function of the flow, and the reanalysis standard deviation is a function of the observational network, the relative information content present in the radiosonde period will vary both spatially and temporally, and will depend on what quantity is under consideration. A complete survey is therefore impossible, but in the next section a brief overview of some commonly used quantities of importance to stratosphere-troposphere interaction is given.

5 Results

Figure 5 shows estimates of the de-seasonalized standard deviation, σ_d, and reanalysis standard deviations σ_s and σ_r for zonal wind in boreal winter and temperature in boreal summer. The standard deviation of the anomaly from the climatology in JRA-55 is used as an estimate of σ_d. The variability of DJF zonal winds is large in the Arctic stratospheric polar vortex, and to a lesser extent in the ~~QBO~~ region of the quasibiennial oscillation (QBO) and on the flanks of the tropospheric jets. The variance of JJA temperatures also shows enhanced variance in the winter stratosphere as well as in the deep tropical stratosphere but the structures are less pronounced. In the troposphere the largest variances are at the poles.

The reanalysis uncertainty is estimated during the satellite period (Fig. 5b) as the variance across six reanalysis products (JRA-55, NCEP-NCAR R1, ERA-40, ERA-Interim, MERRA-2, and CFSR; this choice is further justified below) after first removing ~~the~~ their respective climatological means. The variance is of the order of 0.1 m s^{-1} through much of the extratropics with a slight increase with height, particularly in the winter ~~stratosphere, and~~ upper stratosphere. There is considerably larger inter-reanalysis spread in the deep tropical stratosphere where the lack of strong balance constraints reduces the utility of the thermodynamic measurements available from satellites (Kawatani et al., 2016). Nonetheless the reanalysis uncertainty remains significantly less than the dynamical uncertainty throughout the QBO region, partly due to enhanced dynamical variability, and partly due the observational constraints from radiosondes. In contrast, the inter-reanalysis spread in temperatures is small (0.1 to 0.2 K) throughout most of the summer hemisphere below 10 hPa, but is larger in the upper stratosphere and the winter polar stratosphere. A weak maxima is also seen near the tropical and southern hemisphere tropopauses.

The reanalysis uncertainty during the radiosonde period (Figs. 5ef) is estimated similarly, but using the three full-input reanalyses that cover this period (JRA-55, NCEP-NCAR R1, and ERA-40). Above 10 hPa where data from NCEP-NCAR R1 is not available, the estimate is based on only two products. This results in some weak discontinuities apparent near 10 hPa. The structure of the inter-reanalysis spread is to first order similar to that during the satellite period, but is larger in magnitude. Interhemispheric differences are more apparent, with both wind and temperature spreads in general noticeably larger in the ~~southern hemisphere,~~ Southern Hemisphere (an exception to this is the winds in the upper stratosphere). This is generally consistent with the sparser set of observational constraints. Nonetheless in many regions ~~in~~ it remains substantially smaller than the dynamical variability. Some features with small vertical length-scales are present in the JJA temperature variance, this is likely associated with known artificial vertical temperature oscillations present in ERA-40 (e.g. Manney et al., 2005) (e.g., Randel et al., 2004).

The ‘reanalysis’ uncertainty is, as discussed above, not associated solely with the properties of the observational data available, but also of the assimilation and forecast model used by the respective reanalysis products, and could therefore depend strongly upon which products are included in the calculation. For this reason it is not immediately obvious that the inter-reanalysis spread used here is a reasonable estimate of the reanalysis uncertainty; for instance, certain reanalyses may be outliers for a given quantity and may thus inflate the overall spread.

Figure 6 thus shows pairwise inter-reanalysis differences, computed as a standard deviation over time of the difference between the anomalies from two different reanalyses. For example, if u_i' is the anomalous zonal mean zonal wind of reanalysis i, the difference σ_{ij} between two reanalyses i and j is

$$\sigma_{ij} = \left(\frac{1}{T} \int \left(u_i'(t) - u_j'(t) \right)^2 dt \right)^{1/2}. \quad (4)$$

Entries below the diagonal are computed for the satellite period, those above the diagonal are for the radiosonde period. Entries on the diagonal show the dynamical variability computed from the corresponding reanalysis

$$\sigma_{ii} = \left(\frac{1}{T} \int u_i'(t)^2 dt \right)^{1/2}. \quad (5)$$

The ratio of the inter-reanalysis spread to the dynamical variability (an estimate of α_r and α_s) are indicated by the colour of the off-diagonal cells. Red colours are chosen for ratios greater than 0.3 although this is well below the strict condition of $\alpha < \sqrt{2}$.

Differences are shown for four regions in the winters of the respective hemispheres: (a,b) in the Northern and Southern Hemisphere stratosphere (30 hPa), respectively, and (c,d) in the Northern and Southern Hemisphere troposphere . Note that 100 hPais used to represent the Southern Hemisphere stratosphere while (500 hPa). 30 hPa is used for the Northern Hemisphere stratosphere . This latter is chosen as a representative height for the stratosphere to reduce the effects of the model lid in NCEP-NCAR R1 and NCEP-DOE R2; otherwise the conclusions remain essentially unchanged for 10 hPain the Northern Hemisphere. The estimates of the dynamical variability (along the diagonal) agree closely across all reanalyses, with the exception of 20CR v2 which is significantly less variable in the stratosphere.

The agreement between reanalyses that assimilate some upper air observations In the Northern Hemisphere, the agreement between full-input and conventional-input reanalyses (those other than 20CR v2 and ERA-20C) are in almost all cases below 1030% of the dynamical variability, in both the troposphere and stratosphere. Looking more closely, reanalysis products that share the same or related forecast models tend to be in closer agreement than those from different centres, and there is in general better agreement between the more modern products (JRA-55, ERA-Interim, MERRA-2, CFSR) than between older products. This confirms that the forecast model and assimilation procedure is a contributing factor to the '‘reanalysis’ error. The agreement between In the Northern Hemisphere, the agreement between the conventional-input reanalysis JRA-55C (which does not assimilate satellite observations) and other products is nearly as good as that of JRA-55in the Northern Hemisphere, even in the stratosphere~~, while in the Southern Hemisphere the quality of agreement is degraded; interestingly the agreement in the stratosphere is still higher than with the surface reanalyses, but in the troposphere the latter are in closer agreement.~~

. In the Northern Hemisphere troposphere, the two ~~reanalyses that assimilate only surface observations agree broadly~~ surface-input reanalyses agree with other products to within 30% of the dynamical variability ~~. In the stratosphere and in the southern hemisphere, the differences are considerably larger, but remain smaller than dynamical variability (with the exception of 20CR v2 in the Northern Hemisphere stratosphere)~~in the tropsphere, but this agreement degrades substantially in the stratosphere. Nonetheless, at least for ERA-20C the agreement is to within the dynamical variability, suggesting that surface observations do offer some constraint on the evolution of the stratosphere.

In the Southern Hemisphere the quality of agreement is everywhere weaker than the corresponding cases in the Northern Hemisphere. The full-input reanalyses agree to within 30% in the troppopsphere, and, with a few exceptions, in the stratosphere as well. In the Southern Hemisphere, the conventional-input reanalysis, JRA-55C is more noticeably degraded relative to the agreement between other full-input reanalyses, although the differences are still substantially less than the dynamical variability. The surface-input products also show larger differences in the troposphere.

As expected, differences in the radiosonde era are in general larger than the corresponding differences in the satellite era; the one exception to this is in the Northern Hemisphere stratosphere with 20CR v2, where agreement with JRA-55, ERA-40, and NCEP-NCAR R1 are all apparently slightly improved in the absence of satellite observations. Nonetheless, agreement between these latter ~~three~~ full-input products in the Northern Hemisphere remain very close, showing only a slight degradation within the troposphere, and an agreement between ERA-40 and JRA-55 in the Northern Hemisphere stratosphere to within 10% of the dynamical variability. In contrast, differences in the Southern Hemisphere troposphere approach dynamical variability, and exceed it in the stratosphere.

Given the smaller sample size of products which represent the radiosonde period general conclusions cannot be as strong as those from the satellite period, nonetheless the choice of reanalyses used in Fig. 5 is justified in that no significant outliers are apparent. Lower values of the reanalysis uncertainty would likely be obtained if only more modern reanalyses were included, but this would make comparisons to the radiosonde era impossible. However, given the general improvement in agreement across modern reanalyses seen in the satellite era, it is plausible that further improvements within the radiosonde era are also possible.

Having justified to some extent the estimates of σ_d, σ_r, and σ_s, these can be used to estimate the ratios α_r and α_s, and from ~~there~~ these δ and the effective value of the radiosonde era according to the criteria discussed in the previous section. Following Fig. 5, these quantities are shown for boreal winter zonal winds and austral winter temperatures in Fig. 7.

The ratio α_s is seen to be in general smaller for the zonal winds than for temperatures~~, largely as a result of the larger dynamical variability of the former~~. Consistent with Fig. 5, values are generally smallest in the Northern Hemisphere extra-tropics, below 0.1 for the winds and below 0.2 for temperatures. The ratio is generally below 0.4 for the winds somewhat larger values near the surface in the deep tropics as well as above 10 hPa in the tropics and at high southern latitudes. For ~~the~~ temperatures values are below 0.4 or so in the extratropics up to about 50 hPa, but ~~begin to~~ notably approach 1 near the tropopause in the tropics ~~and~~ where dynamical variability is small, as well as in the Southern Hemisphere, and through much of the stratosphere.

The ratio α_r shares many of the structural features present in α_s but with generally larger values. Most importantly for the present discussion, the Northern Hemisphere extratropical winds show values still in general below 0.2~~, although these values approach~~. For zonal winds, the ratio exceeds 0.5 but remains below 1 ~~in~~ through most of the Southern Hemisphere~~and stratosphere. Again~~, indicating the observations are less effective at constraining the winds in this hemisphere, but there is still some information common across reanalyses. As with α_s, α_r ~~for temperatures are larger, in particular near the tropopause.~~ is larger for temperatures than for zonal winds, particularly near the tropical and Southern Hemisphere tropopause where values are well above 1. Values in the Northern Hemisphere extratropics ~~remain small.~~ through the lower stratosphere remain small, but the summertime mid-stratospheric temperatures (where dynamical variability is relatively weak) are not well constrained. Much of the wintertime Southern Hemisphere also shows values near 1.

~~These ratios suggest that~~ Using these values of α_r and α_s, Fig. 5ef show the calculated value of δ. The values for the zonal wind remains ~~in fact~~ quite close to 1 through the Northern Hemisphere and tropics ~~, and are in fact only somewhat reduced for~~ in boreal winter. In the Southern Hemisphere below 10 hPa ~~. Despite considerable additional uncertainty , this suggests that JJA winds are still well-enough constrained by observations~~ the values are reduced, but perhaps surprisingly remain above 0.5. This reflects to some extent the fact that the underlying reanalysis uncertainty σ_s is larger in Southern Hemisphere than in the Northern Hemisphere, even during the satellite era. These values suggest that DJF winds are constrained well-enough by observations in the radiosonde era that they may be of some value ~~. Although not shown here, this is true also of DJF winds~~towards reducing uncertainty. This is, however, not the case for ~~DJF~~ JJA temperatures in the Southern Hemisphere (Fig. 5f, or in fact ~~JJA~~ for JJA winds or DJF temperatures, though ~~this is not shown~~ these latter cases are not shown explicitly), for which values of δ are in many cases below 0; this is notably the case for temperatures near the tropical tropopause as well.

In ~~practice~~ summary, these criteria show clear value in including the radiosonde era in dynamical analyses of Northern Hemisphere quantities from the troposphere up to the mid-stratosphere. There is a possible suggestion that useful information may be gained for winds in the Southern Hemisphere summer winds as well. On the other hand, for much of the rest of the Southern Hemisphere quantities this is not the case. Temperatures near the tropical tropopause also show significantly worse agreement during the radiosonde period.

As they are based on the overall variance, these estimates are most sensitive to the dominant dynamical structures of interannual variability in the flow, which have typically relatively longer time scales and larger length scales. These bulk estimates may not therefore imply that the observational constraints on dynamical processes at shorter timescales are equally strong. To begin to assess this point, Fig. 8 compares the power spectra of deseasonalized winds from JRA-55 in the stratosphere and troposphere with the power spectra of pairwise differences between JRA-55 and other reanalyses. ~~The~~ These provide frequency-dependent estimates of σ_d and σ_o, respectively, and thus the ratio of these two spectra in the corresponding eras ~~can thus be used as~~ provides a frequency-dependent estimate of α_s^2 and α_r^2. Such spectra are shown for Northern Hemisphere winds in the stratosphere (Fig. 8a,b) and in the troposphere (Fig. 8c,d).

In all cases the raw spectrum of JRA-55 is shown as a reference; curves for all other reanalyses show the power spectrum of the differences between those reanalyses and JRA-55. During satellite era differences from most reanalyses at low frequencies are almost two two to three orders of magnitude smaller than the spectrum(, consistent with the 5-10% estimate of the raw difference) differences since these plots show the variance instead of the standard deviation. However, fluctuations These values can be compared to the horizontal line shown at a value of 2, expected if observations are providing no constraint on the flow. Fluctuations at higher frequencies reach the same order as the dynamical variability at timescales of a few days in the stratosphere; in the troposphere differences amongst the more modern reanalyses remain below dynamical variability down to the highest frequency considered (corresponding to a period of 6 hours). Within the stratosphere differences from NCEP-NCAR R1 and NCEP-DOE R2 are significantly larger than other reanalyses at all frequencies and the differences from ERA-20C and 20CR v2 are as large as of the order of the reference spectrum. Within the troposphere the surface surface-input reanalyses are still noticeably in less good weaker agreement with JRA-55, with difference spectra that approach the reference spectra at frequencies corresponding to periods less than half a week or so.

During the radiosonde period era (Fig.8b,d) the differences are, as expected, larger than during the satellite periodera, although similar features can be noted with better agreement between JRA-55 and ERA-40, and significantly worse agreement with the surface reanalyses.

surface-input reanalyses. This suggests that processes with timescales even as short as a few days are still significantly constrained in the Northern Hemisphere extratropics, although this constraint is not as strong (relative to dynamical variability) as is the case for processes on timescales of a month or longer.

A similar spectral analysis could be applied spatially to determine which spatial scales which are reliable. However this has not been directly considered and would be better applied to fully three dimensional data as opposed to the zonal means considered here.

Up to this point the analysis has considered both the radiosonde and satellite eras to be to some extent uniform in time in their properties; of course yet the observational record evolved during these periods as well. To consider briefly the evolution of the observational constraint over time, the ratio α_r α can be estimated for each month individuallyover time; in . In this case we take consider pairwise differences between JRA-55 and other reanalyses as an estimate of σ_o, and the standard deviation of JRA-55 itself as an estimate of σ_d. In all cases the time-series are first de-seasonalized. These ratios can then be used to estimate δ; however, to do so one must assume an appropriate reference value for α_s, here taken to be 0.1 which is roughly appropriate for both quantities based on Fig. 7a. For sufficiently high values of α_r the estimate will also depend on β; in this regime a time-dependent value of δ is not strictly meaningful, although a small value can still be considered indicative of diminished value.

Since the interest is primarily in the early part of the record, Fig. 9 shows this ratio for zonal winds at in the Northern Hemisphere stratosphere (at 60 N, 30 hPa), and in the Southern Hemisphere troposphere (at 45 S, 500 hPa), spanning from 1958 through 1986. The month by month values fluctuate considerably, but show nonetheless a distinct annual cycle with higher values of δ lower values of α during the respective winter months when the dynamical variability is higher. A clearer trend can be observed by considering δ computed from 12-month running averages of α (bold lines in Fig.9, which suggests

~~that the value of the full-input reanalyses remains high through essentially~~). In the Northern Hemisphere stratosphere, Values for ERA-40 remain well below 0.5 through nearly all of the ~~radiosonde era in the Northern Hemisphere stratosphere, while in the Southern Hemisphere troposphere the value diminishes rapidly prior to 1979. The reasonably good agreement across full-input reanalyses in the Northern Hemisphere stratosphere even towards the beginning of the time period considered here suggests that even the 1950s may be of interest, however, of all full-input reanalyses considered, only~~ period in question, and NCEP-NCAR R1 ~~includes this decade.~~

~~The surface reanalyses, in particular~~ is only somewhat larger. Although the methodology used here cannot yet be used to examine the period prior to 1958, these relatively low values suggest that even earlier periods could be of value. This speculation is supported by the results of Hersbach et al. (2017) who found this period to be of value in particular for constraining the evolution of the QBO.

The surface-input reanalyses show large fluctuations over time, but less of a clear trend. For ERA-20C the value of α remains close to 1 through much of the period, though at the beginning of the period the value is only slighly larger than for NCEP-NCAR R1. The values for 20CR v2 ~~, are nearly as good as the full-input reanalyses in~~ are systematicaly larger, not far below the limit of $\sqrt{2}$ despite the lower overall variance at these heights seen in Fig. 6.

In the Southern Hemisphere ~~troposphere, but their value in the Northern Hemisphere stratosphere is substantially less than those of the full-input reanalyses~~again values show a clear seasonal cycle; while there are times of the year during which the agreement is better, the 12-month runing average are above 1 for all products through the 1960s, dropping somewhat through the early 1970s and to values of less than 0.5 only after 1979. This suggests that the tropospheric flow is only weakly constrained by the observations prior to 1979. In this case the 20CR v2 shows somewhat better agreement with JRA-55 than ERA-20C through the early 1980s.

The assessment of inter-reanalysis differences presented here suggest that there is considerable value for dynamical studies in including the radiosonde era, particularly in the extratropical Northern Hemisphere. The criteria discussed suggest that for lower-frequency, large-scale processes such as those responsible for stratosphere-troposphere coupling during stratospheric sudden warmings, including the radiosonde era could ~~reduced~~ reduce confidence intervals by close to 20%, despite the increase in reanalysis uncertainty during this time. To assess whether this is in fact the case, Fig. 10 presents bootstrap estimates of uncertainties (at the 95% level) on composites of several dynamical quantities fundamental to this coupling: the vertically integrated zonal wind, vertically integrated meridional momentum fluxes, and meridional heat fluxes at 100 hPa. The vertical integral is taken from 1000 hPa to 100 hPa (see, e.g., Hitchcock and Simpson, 2016). The bootstrap estimates are carried out by generating a large number of synthetic composites by selecting N events with replacement from the full period (shown in solid lines with shaded confidence intervals), and from the satellite period (shown in dashed lines with outlined confidence intervals).

Importantly, any systematic error present in these quantities during the radiosonde era will contribute to the bootstrapped confidence intervals. The fact then that in each case confidence intervals are (with some regional exceptions; not shown explicitly) reduced by on the order of 20% ~~(not shown explicitly) confirms~~ suggests that any such systematic errors are small relative to the sampling error.

As was the case with the event frequencies shown in Fig. 3, the composite means agree nearly everywhere to within estimated confidence intervals, as should be the case. Within these uncertainties, the tropospheric jet shift is seen at somewhat lower latitudes during the full period with a less pronounced low-latitude signal; the momentum flux anomalies are somewhat more positive, and the heat-flux anomalies during the recovery phase suggest somewhat more suppression of the upward wave flux. While the differences in composite means are modest, including this period reduces the confidence intervals on these quantities by the expected amount, providing better observational constraints on dynamical understanding and modeling efforts.

6 Conclusions

The growth of satellite observations providing global coverage following 1979 The advent of more advanced satellite-based sounding instruments in the late 1970's resulted in major improvements in the monitoring of the detailed state of the atmosphere. However, Nonetheless, 'conventional' upper-air observations play an important complementary role, and the network of surface and radiosonde observations in the period from 1958 to 1978 were remarkably effective in constraining many features of the general circulation, even in the boreal lower stratosphereplace prior to this period represent a valuable resource for observationally constraining atmospheric variability. For dynamical studies that rely on statistical composites of specific anomalous conditions, the dominant source of error is in many cases that of sampling variability, and in this context the radiosonde period represents a valuable extension of the observational record, allowing in principle arises from sampling this atmospheric variability, not from observational uncertainties.

In particular, this study has considered the value of the 'radiosonde' era from 1958 to 1978 relative to the 'satellite' era from 1979 to 2010, using differences between presently available reanalysis products to characterize the constraint provided by the observations in these two periods. In principal, including the radiosonde era allows for up to a reduction of 20% in confidence intervals associated with the dynamical variability.

The value of this record the radiosonde era towards reducing the overall sampling uncertainty in composites has been quantified is quantified by equation (3). This depends on the ratio of the 'reanalysis' uncertainty (including errors uncertainty arising both from the precision of the underlying observations as well as that arising from the assimilation process) to the dynamical uncertainty (the variability of the dynamical phenomena themselves). In general this depends also on the relative length of A key conclusion to draw from this relationship is that even if the reanalysis uncertainty is significantly greater in the radiosonde era to the total time period considered, but when the dynamical variability dominatesthe overall uncertainty, this dependency drops out. than in the satellite era, so long as the dynamical uncertainty dominates, the radiosonde era will be of nearly equivalent value to the satellite era. However, since this criterion assesses the relative value of the two periods, it is important as well to consider directly the ratio of the reanalysis uncertainty to the dynamical uncertainty. If this is too large, this indicates a more significant influence of the underlyin forecast model.

Since this quantity is in practice a function of both Since these criteria depend on the physical properties of the climate system, the observations available, and of the reanalysis forecast model and assimilation system, this criteria they must be applied on a case-by-case basis, and the . The present work cannot hope to provide a comprehensive survey. However, basic zonal mean quantities including zonal winds, temperatures, and fluxes of momentum and heat, as archived for 12 reanalysis products (see Table 1) by Martineau (2017), have been considered here.

For all quantities considered, the reanalysis uncertainty in the Northern Hemisphere extratropics from the surface up to the mid-stratosphere (about 10 hPa) is found to be sufficiently small relative to the dynamical variability to make the radiosonde era of clear value in reducing composite uncertainties. For zonal mean zonal winds, the interannual variability is such that despite larger reanalysis uncertainties, this is also the case for tropical winds (even in the stratosphere) and even Southern Hemisphere winds ~~are of potential value . Because the dynamical variability of temperature is smaller, the reanalysis uncertainty in the radiosonde era is relatively large and suggests that~~ may be of some value in the austral summer. However, temperatures through much of the Southern Hemisphere ~~is~~ are not well enough constrained to be worth including the radiosonde era. This is also notably true of temperatures in the tropical tropopause layer.

This test has also ~~be~~ been applied to the ~~surface~~ surface-input reanalyses ERA20c and 20CR v2. The statistics of differences between these products and full-input reanalyses clearly indicate that, at least for ERA20c, their stratospheric evolution bears some meaningful resemblance to reality. However, ~~the test indicates that, relative to the constraint available from~~ this constraint is still much weaker to that available to full-input ~~reanalyses during the satellite era, their errors are too large to meaningfully constrain dynamical variability(see Fig. 9)~~or even conventional-input products, with inter-reanalysis differences of similar magnitude to the dynamical variability. Furthermore, while differences between other reanalyses are reduced when considering fixed dates for stratospheric sudden warmings, for the ~~surface~~ surface-input reanalyses the comparison is improved when considering per-reanalysis dates, suggesting that, in these ~~surface~~ surface-input reanalyses, stratospheric sudden warmings are ~~more~~ at least as much a product of the forecast model dynamics than a result of assimilated observations.

While ~~this~~ these criteria does not consider the possibility of systematic biases in the radiosonde era, direct bootstrap estimates generally confirm this reduction in uncertainty of several dynamical quantities relevant to stratosphere-troposphere coupling following stratospheric sudden warmings in the Northern Hemisphere. ~~These estimates are sensitive to systematic biases (at least any relative to those in the satellite era), suggesting that any such biases are negligible for these quantities.~~

~~Finally~~As a final note, while considerable improvements have been documented for more modern reanalyses during the satellite period ~~(e.g. Long et al., 2017)~~(e.g., Long et al., 2017), there are at present not enough modern reanalyses that cover the radiosonde era to clearly document improvements over this earlier period. ~~Nonetheless, it~~ It seems likely that ~~further attention on this period could produce further~~ similar attention on the radiosonde era could produce similar improvements. Given the value of this period for dynamical studies ~~, such attention from the reanalyses centres would be~~ demonstrated in this and other recent studies (Hersbach et al., 2017; Gerber and Martineau, 2018), the intent to include this period in two upcoming products (ERA-5 and JRA-3Q) is welcome.

7 Data availability

All analysis is based on the zonal mean dataset kindly provided by Patrick Martineau which is available online from the Centre for Environmental Data Analysis Martineau (2017).

Acknowledgements. The author thanks Sean Davis and Gloria Manney for helpful discussions, as well the lead authors of the SRIP chapter on Stratosphere-Troposphere Coupling, Patrick Martineau and Ed Gerber, for their support of this work. The reviewer comments of Adrian Simmons, Ed Gerber, and an anonymous referee lead to significant improvements in the text, and were also much appreciated.

References

Birner, T. and Albers, J. R.: Sudden Stratospheric Warmings and Anomalous Upward Wave Activity Flux, Sci. Online Lett. Atmos., 13A, 8–12, https://doi.org/10.2151/sola.13A-002, 2017.

Charlton, A. J. and Polvani, L. M.: A new look at stratospheric sudden warmings. Part I: Climatology and modelling benchmarks, J. Clim., 20, 449–469, 2007.

Charlton-Perez, A. J., Baldwin, M. P., Birner, T., Black, R. X., Butler, A. H., Calvo, N., Davis, N. A., Gerber, E. P., Gillett, N., Hardiman, S., Kim, J., Krüger, K., Lee, Y.-Y., Manzini, E., McDaniel, B. A., Polvani, L., Reichler, T., Shaw, T. A., Sigmond, M., Son, S.-W., Toohey, 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, Q. J. R. Meteorol. Soc., 137, 1–28, https://doi.org/10.1002/qj.776, 2011.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., and Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, Q. J. R. Meteorol. Soc., 137, 553–597, https://doi.org/10.1002/qj.828, 2011.

Deser, C., Simpson, I. R., McKinnon, K. A., and Phillips, A. S.: The Northern Hemisphere extra-tropical atmospheric circulation response to ENSO: How well do we know it and how do we evaluate models accordingly?, J. Clim., 30, 5059–5082, https://doi.org/10.1175/JCLI-D-16-0844.1, 2017.

Dunn-Sigouin, E. and Shaw, T. A.: Comparing and contrasting extreme stratospheric events, including their coupling to the tropospheric circulation, J. Geophys. Res., 120, 1374–1390, https://doi.org/10.1002/2014JD022116, 2014.

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.

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. Clim., 30, 5419–5454, https://doi.org/10.1175/JCLI-D-16-0758.1, 2017.

Gerber, E. P. and Martineau, P.: Quantifying the variability of the annular modes: Reanalysis uncertainty vs. sampling uncertainty, Atmos. Chem. Phys., 18, 17 099–17 117, https://doi.org/10.5194/acp-18-17099-2018, 2018.

Gómez-Escola, M., Fueglistaler, S., Calvo, N., and Barriopedro, D.: Changes in polar stratospheric temperature climatology in relation to stratospheric sudden warming occurrence, Geophys. Res. Lett., 39, L22 802, https://doi.org/10.1029/2012GL053632, 2012.

Hersbach, H., Brönnimann, S., Haimberger, L., Mayer, M., Villiger, L., Comeaux, J., Simmons, A., Dee, D., Jourdain, S., Peubey, C., Poli, P., Rayner, N., Sterin, A. M., Stickler, A., Valente, M. A., and Worley, S. J.: The potential value of early (1939–1967) upper-air data in atmospheric climate reanalysis, Q. J. R. Meteorol. Soc., 143, 1197–1210, https://doi.org/10.1002/qj.3040, 2017.

Hitchcock, P. and Simpson, I. R.: The downward influence of stratospheric sudden warmings, J. Atmos. Sci., 71, 3856–3876, https://doi.org/10.1175/JAS-D-14-0012.1, 2014.

Hitchcock, P. and Simpson, I. R.: Quantifying forcings and feedbacks following stratospheric sudden warmings, J. Atmos. Sci., 73, 3641–3657, https://doi.org/10.1175/JAS-D-16-0056.1, 2016.

Hitchcock, P., Shepherd, T. G., and Manney, G. L.: Statistical characterization of Arctic Polar-night Jet Oscillation events, J. Clim., 26, 2096–2116, https://doi.org/10.1175/JCLI-D-12-00202.1, 2013.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Leetmaa, A., Reynolds, R., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K. C., Ropelewski, C., Wang, J., Jenne, R., and Joseph, D.: The NCEP/NCAR 40-year reanalysis project, Bull. Amer. 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. Amer. Meteor. Soc., 83, 1631–1643, 2002.

Kawatani, Y., Hamilton, K., Miyazaki, K., Fujiwara, M., and Anstey, J. A.: Representation of the tropical stratospheric zonal wind in global atmospheric reanalyses, Atmos. Chem. Phys., 83, 1631–1643, https://doi.org/10.5194/acp-16-6681-2016, 2016.

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. Meteor. Soc. Japan, 93, 5–48, https://doi.org/10.2151/jmsj.2015-001, 2015.

Kodera, K., Mukougawa, H., Maury, P., Ueda, M., and Claud, C.: Absorbing and reflecting sudden stratospheric warming events and their relationship with tropospheric circulation, J. Geophys. Res., 121, 80–94, https://doi.org/10.1002/ 2015JD023359, 2015.

Labitzke, K.: Interannual variability of the winter stratosphere in the Northern Hemisphere, Mon. Wea. Rev., 105, 762–770, 1977.

Lehtonen, I. and Karpechko, A. Y.: Observed and modeled tropospheric cold anomalies associated with sudden stratospheric warmings, J. Geophys. Res., 121, 1591–1610, https://doi.org/10.1002/2015JD023860, 2016.

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.

Manney, G. L., Krüger, K., Sabutis, J. L., Sena, S. A., and Pawson, S.: The remarkable 2003-2004 winter and other recent warm winters in the Arctic stratosphere since the late 1990s, J. Geophys. Res., 110, D04 107, https://doi.org/10.1029/2004JD005367, 2005.

Manzini, E., Karpechko, A. Y., Anstey, J., Baldwin, M. P., Black, R. X., Cagnazzo, C., Calvo, N., Charlton-Perez, A., Christiansen, B., Davini, P., Gerber, E., Giorgetta, M., Gray, L., Hardiman, S. C., Lee, Y.-Y., Marsh, D. R., McDaniel, B. A., Purich, A., Scaife, A. A., Shindell, D., Son, S.-W., Watanabe, S., and Zappa, G.: Northern winter climate change: Assessment of uncertainty in CMIP5 projections related to stratosphere-troposphere coupling, J. Geophys. Res., 119, 7979–7998, https://doi.org/10.1002/2013JD021403, 2014.

Martineau, P.: Zonal-mean dynamical variables of global atmospheric reanalyses on pressure levels, https://doi.org/10.5285/b241a7f536a244749662360bd7839312, 2017.

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, 2018.

Matsuno, T.: A Dynamical model of the stratospheric sudden warming, J. Atmos. Sci., 28, 1479–1494, 1971.

McIntyre, M. E.: How well do we understand the dynamics of stratospheric warmings?, J. Meteor. Soc. Japan, 60, 37–65, 1982.

Mitchell, D. M., Gray, L. J., Anstey, J., Baldwin, M. P., and Charlton-Perez, A. J.: The Influence of Stratospheric Vortex Displacements and Splits on Surface Climate, J. Clim., 26, 2668–2682, https://doi.org/10.1175/JCLI-D-12-00030.1, 2013.

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. Meteor. Soc. Japan, 85, 369–432, https://doi.org/10.2151/jmsj.85.369, 2007.

Poli, P., Hersbach, H., Tan, D., Dee, D., Thépaut, J.-N., Simmons, A., Peubey, C., Laloyaux, P., Komori, T., Berrisford, P., Dragani, R., Trémolet, Y., Holm, E., Bonavita, M., Isaksen, L., and Fisher, M.: The Data Assimilation System and Initial Performance Evaluation of the ECMWF Pilot Reanalysis of the 20th Century Assimilating Surface Observations Only (ERA-20C), Tech. Rep. 14, ECMWF, Reading, UK, 2013.

Randel, W., Udelhofen, P., Fleming, E., Geller, M., Gelman, M., Hamilton, K., Karoly, D., Ortland, D., Pawson, S., Swinbank, R., Wu, F., Baldwin, M., Chanin, M.-L., Keckhut, P., Labitzke, K., Remsberg, E., Simmons, A., and Wu, D.: The SPARC Intercomparison of Middle-Atmosphere Climatologies, J. Clim., 17, 986–1003, https://doi.org/10.1175/1520-0442(2004)017<0986:TSIOMC>2.0.CO;2, 2004.

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

Saha, S., Moorthi, S., Pan, H.-L., Wu, X., Wang, J., Nadiga, S., Tripp, P., Kistler, R., Woollen, J., Behringer, D., Liu, H., Stokes, D., Grumbine, R., Gayno, G., Wang, J., Hou, Y.-T., Chuang, H.-Y., Juang, H.-M. H., Sela, J., Iredell, M., Treadon, R., Kleist, D., van Delst, P., 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. Amer. Meteor. Soc., 91, 1015–1057, https://doi.org/10.1175/2010BAMS3001.1, 2010.

Scherhag, R.: Die explosionsartigen Stratosphärenerwärmungen des Spätwinters, Ber. Dtsch. Wetterdienst (US Zone), 6, 51–63, 1952.

Sigmond, M., Scinocca, J. F., Kharin, V. V., and Shepherd, T. G.: Enhanced seasonal forecast skill following stratospheric sudden warmings, Nat. Geosci., 6, 98–102, https://doi.org/10.1038/NGEO1698, 2013.

Simmons, A., Hortal, M., Kelly, G., McNally, A., Untch, A., and Uppala, S.: ECMWF Analyses and Forecasts of Stratospheric Winter Polar Vortex Breakup: September 2002 in the Southern Hemisphere and Related Events, J. Atmos. Sci., 62, 668–689, https://doi.org/10.1175/JAS-3322.1, 2005.

Simpson, I. R., Hitchcock, P., Seager, R., and Wu, Y.: The downward influence of uncertainty in the Northern Hemisphere stratospheric polar vortex response to climate change, J. Clim., 31, 6371–6391, https://doi.org/10.1175/JCLI-D-18-0041.1, 2018.

Taguchi, M.: A study of different frequencies of major stratospheric sudden warmings in CMIP5 historical simulations, J. Geophys. Res., 122, 5144–5156, https://doi.org/10.1002/2016JD025826, 2017.

Uppala, S. M., Kållberg, P. W., Simmons, A. J., Andrae, U., Bechtold, V. D. C., 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., Berg, L. V. D., Bidlot, J., Bormann, N., Caires, S., Chevallier, F., Dethof, A., Dragosavac, M., Fisher, M., Fuentes, M., Hagemann, S., Hólm, 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, Q. J. R. Meteorol. Soc., 131, 2961–3012, https://doi.org/10.1256/qj.04.176, 2005.

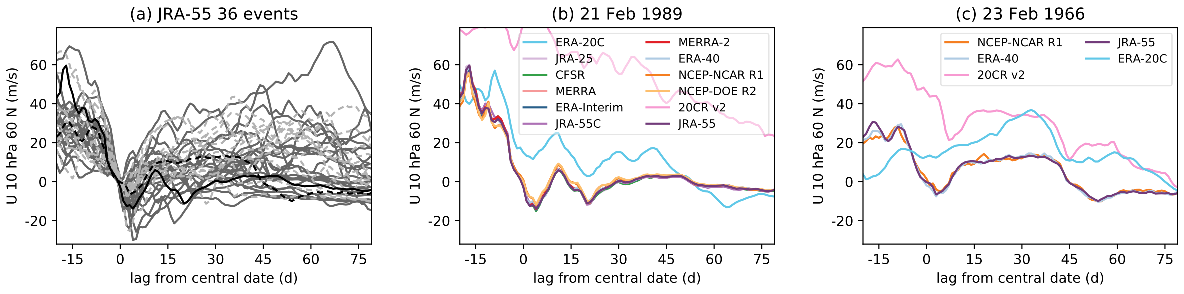

Figure 1. (a) Winds from JRA-55 for 36 sudden warmings. Events from the satellite period are in dark grey, those from the radiosonde period are in light grey and are dashed. (b) Winds for a single satellite-period event for all reanalyses; this event is shown by the black line in (a). (c) Winds for a single radiosonde-period event for all reanalyses covering this period; this event is shown by the dashed black line in (a).

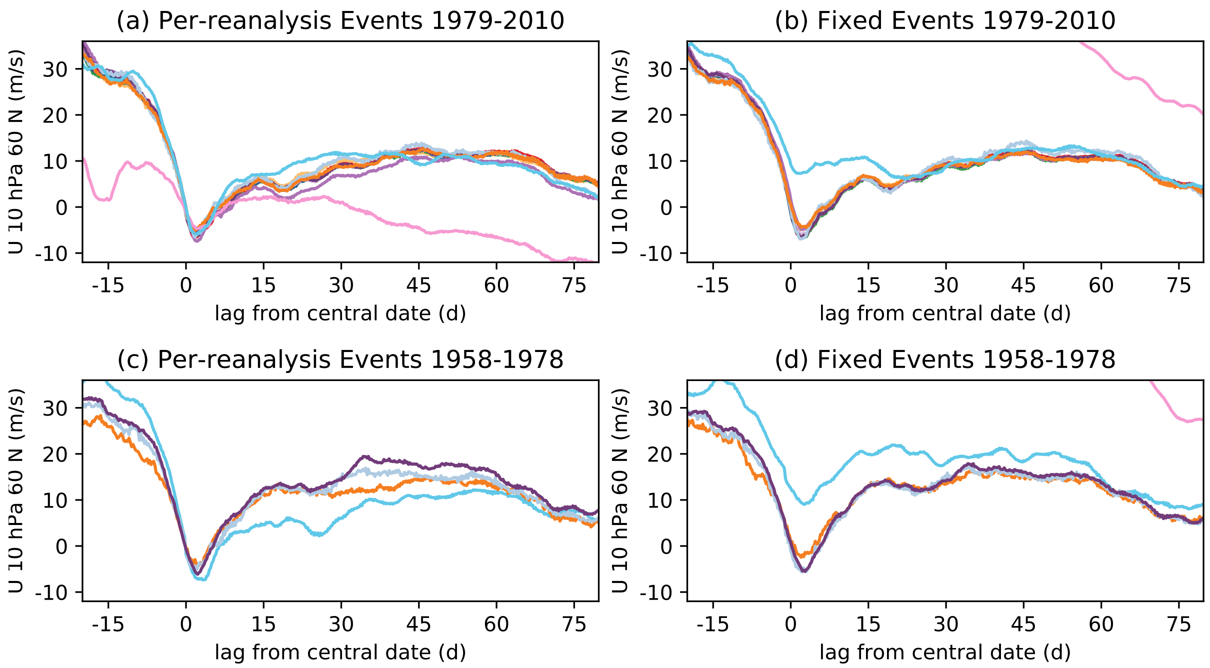

Figure 2. Composites of zonal mean zonal wind at 10 hPa, 60° N during stratospheric sudden warmings for events during the satellite era (a,b) and the radiosonde era (c,d). Events in (a,c) are determined by applying the wind reversal criteria of Charlton and Polvani (2007) to each reanalysis individually, while those in (b,d) are taken to be common across all reanalyses. Line colours are as in Fig. 1.

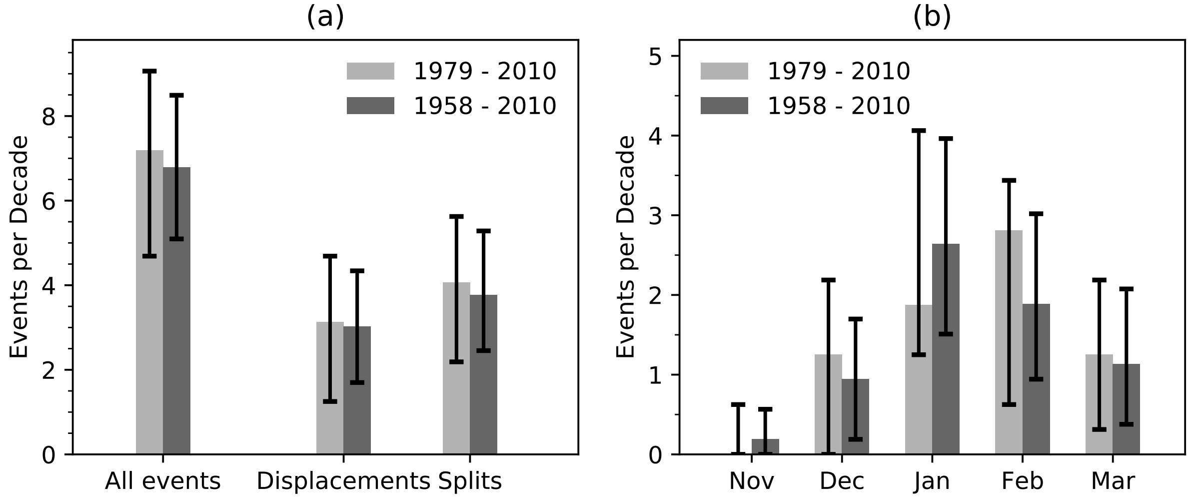

Figure 3. (a) Frequency of all events, and of events classified as splits or displacements for details for the satellite period versus for the radiosonde period. (b) Same as (a) but for each month of extended winter. Error bars indicate 95% confidence intervals, see text for details.

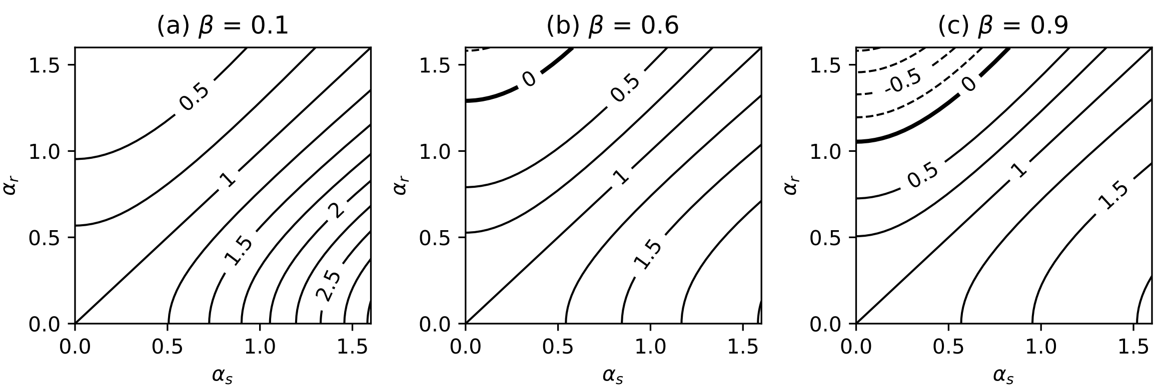

Figure 4. The effective value δ of radiosonde-era degrees of freedom relative to that of satellite-era degrees of freedom in reducing the overall uncertainty. Shown as a function of α_r and α_s for three values of β: (a) 0.1 (radiosonde era much longer than satellite era), (b) 0.6 (roughly appropriate for the observational records considered here) and (c) 0.9 (radiosonde era much shorter than satellite era). Contour interval is ~~0.5~~0.25, with the 0 contour indicated ~~by the~~ in bold~~line~~.

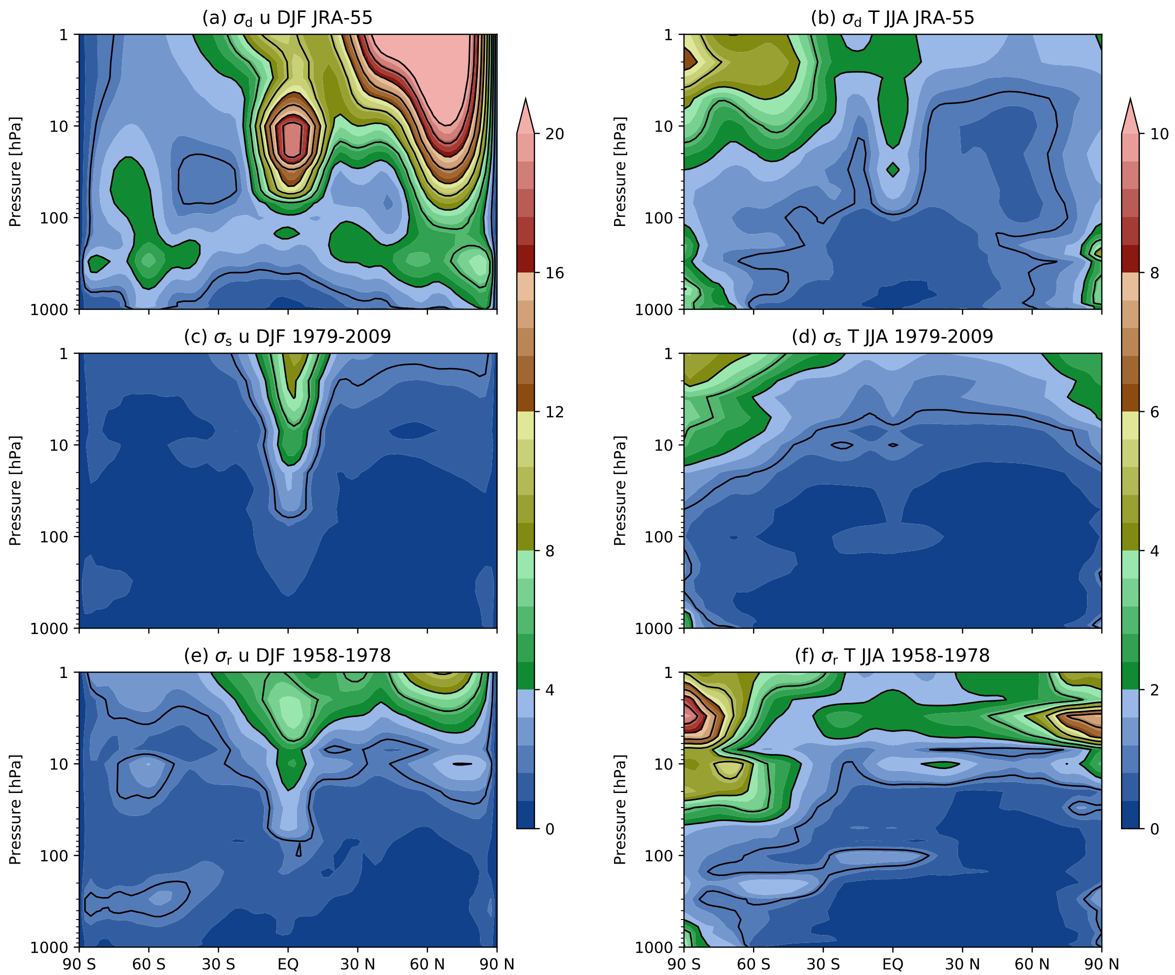

Figure 5. Standard deviation of de-seasonalized (a) winds in DJF and (b) temperatures in JJA from the JRA-55 reanalysis over the satellite period. (c,d) Standard deviation of the differences in same quantities (respectively) across six reanalysis products for the satellite period. (e,f) As in (c,d) but across three reanalysis products for the radiosonde period. See text for details.

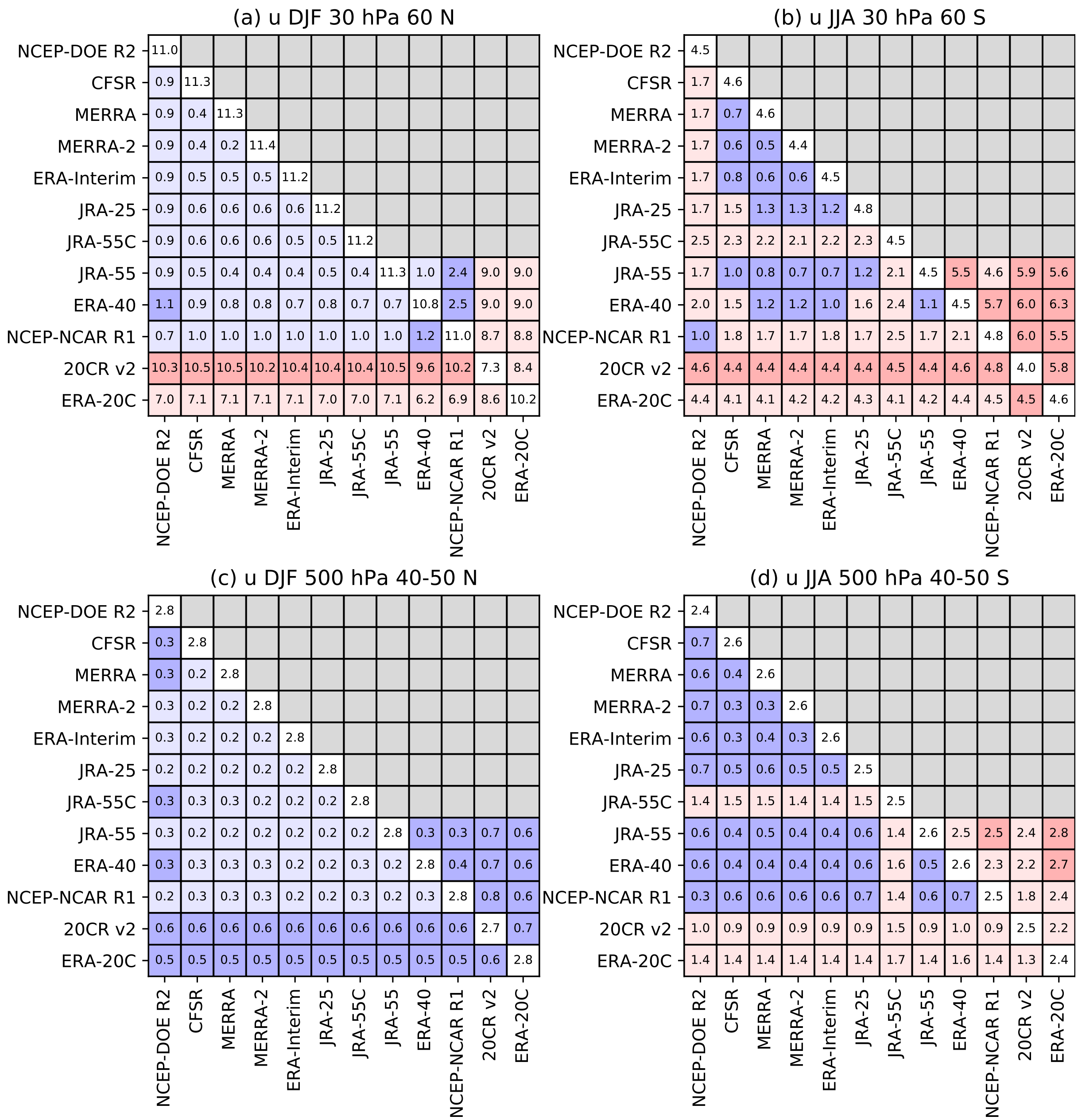

Figure 6. Standard deviations of pair-wise differences between winds in different reanalysis products at (a) 30 hPa, 60° N (DJF), (b) 100 hPa, 60° S (JJA), (c) 500 hPa, 40-50° N (DJF), and (d) 500 hPa, 40-50° S (JJA). All quantities are in m s^{-1}. The diagonal elements show the de-seasonalized standard deviation of the corresponding quantity, elements below the diagonal show differences for the satellite era, and elements above the diagonal show differences for the radiosonde era. Elements are shaded by the ratio of the difference to the mean of the dynamical standard deviations from the corresponding two diagonal elements; light blue (less than 10%), dark blue (10% to 30%), light red (30% to 100%), and dark red (greater than 100%).

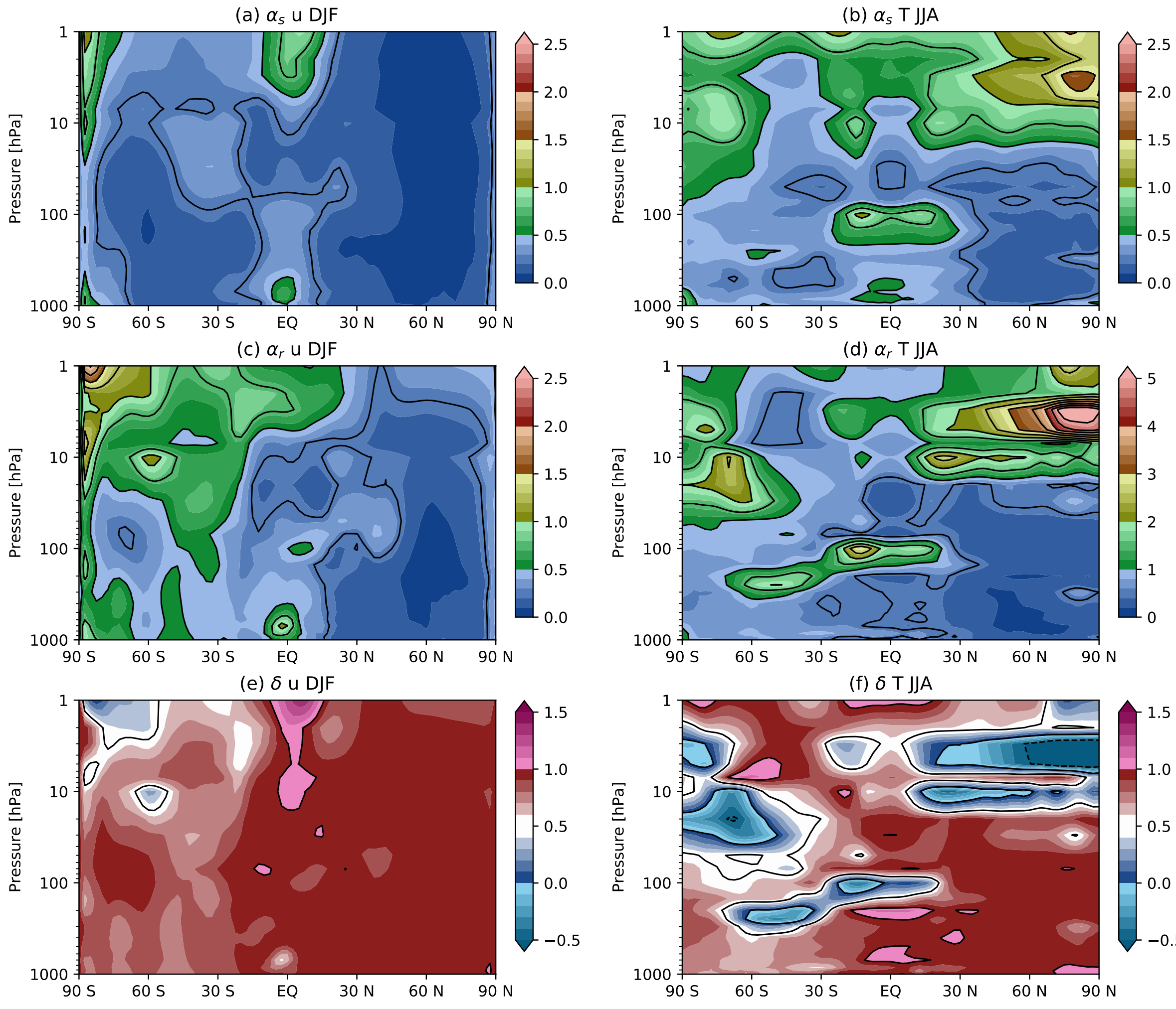

Figure 7. Ratios (a,~~c~~b) ~~γ~~ α_s and (~~b~~c,d) ~~α~~ α_r, and (e,f) the effective value δ of radiosone-era degrees of freedom as defined in Section 3 for (a,~~b~~c,e) zonal winds in DJF and (~~c~~b,d,f) temperatures in JJA. Note the different scale for panel (d).

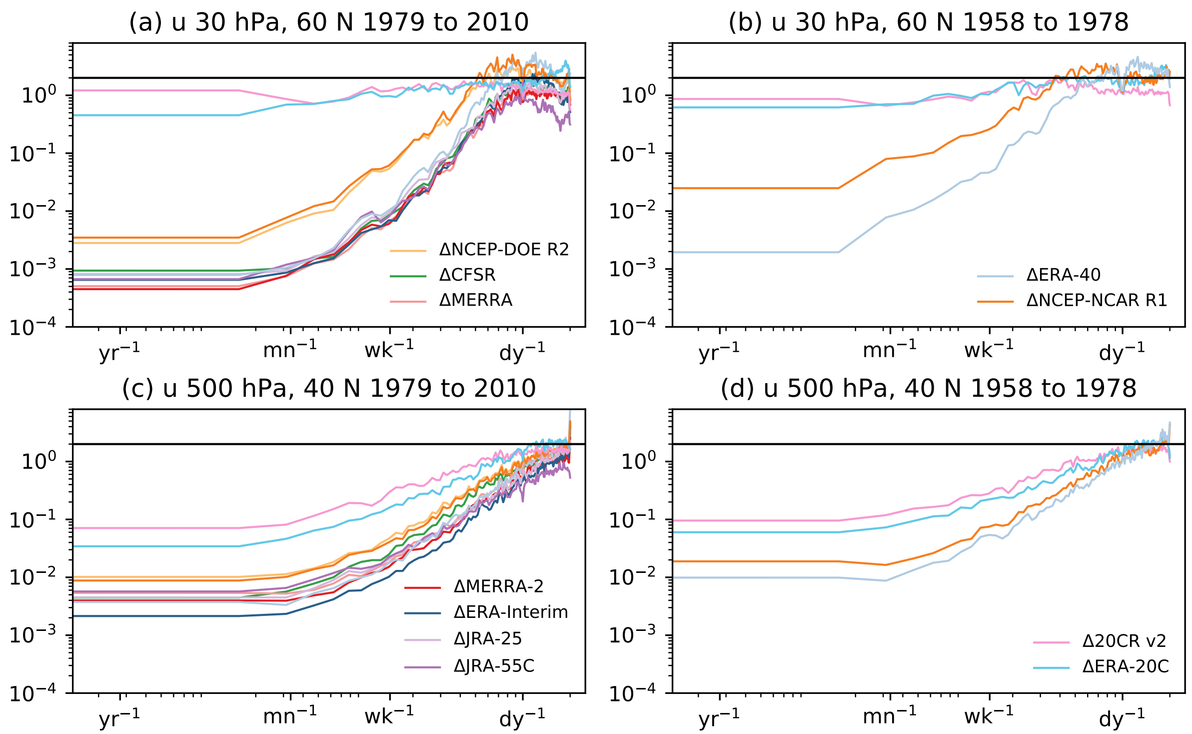

Figure 8. Power spectrum Ratio of winds in JRA-55 and the power spectrum of the differences in zonal winds between JRA-55 and other reanalyses (as indicated in the legend), and the power spectrum of winds in JRA-55 at itself. Winds are deseasonalized and from (a,b) 30 hPa, 60 N and (c,d) 500 hPa, 40 N in the radiosonde and satellite periods, respectivelyera (left panels) and radiosonde era (right panels). Note that the legend is divided across the panels but applies equally to each. Frequencies corresponding to periods of one year, one month (30 days), one week, and one day are indicated on the horizontal axis. The black horizontal line is at 2, indicative of the lack of observational constraints (see text).

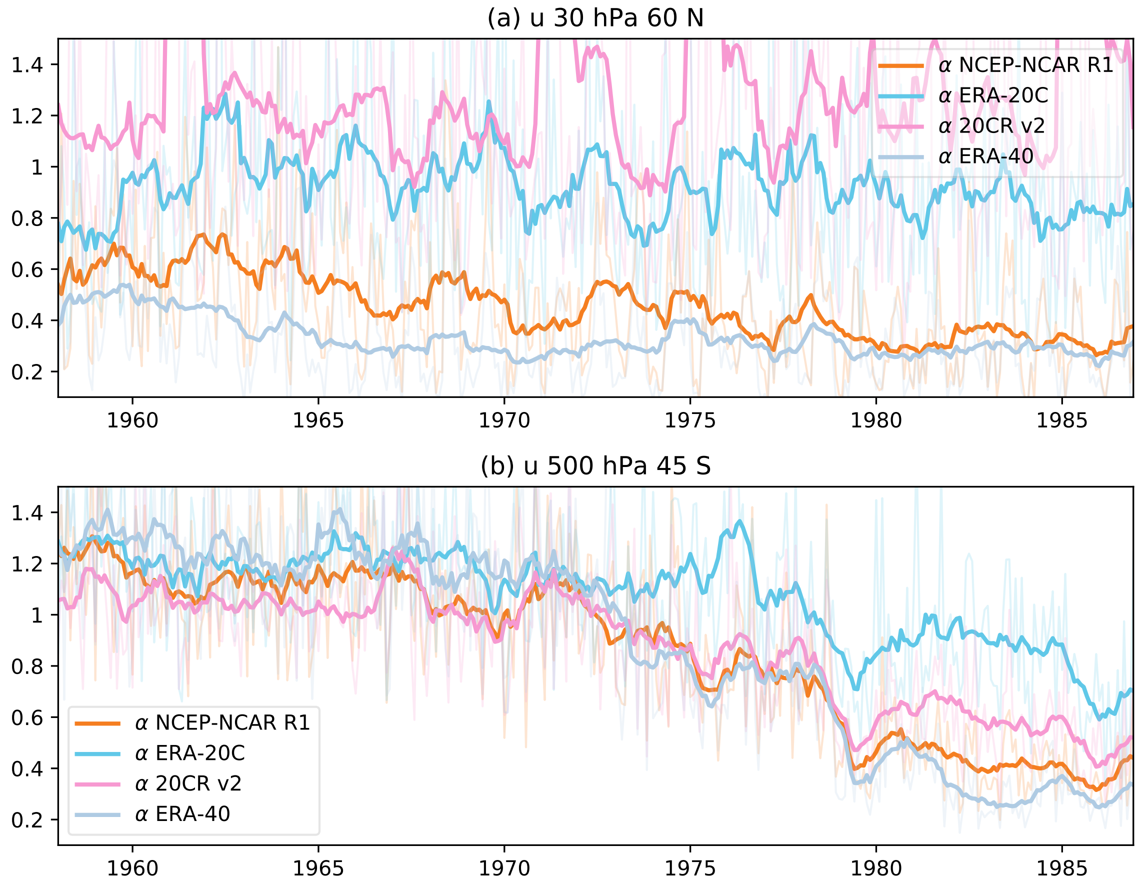

Figure 9. Time-dependent estimate of δ α for (a) U at 30 hPa, 60° N and (b) U at 500 hPa, 45° S. The faint lines are computed based on month-by-month estimates of α_r variability (see text for details), while bold lines are computed based 12-month running means of α_r. See text for detailsα.

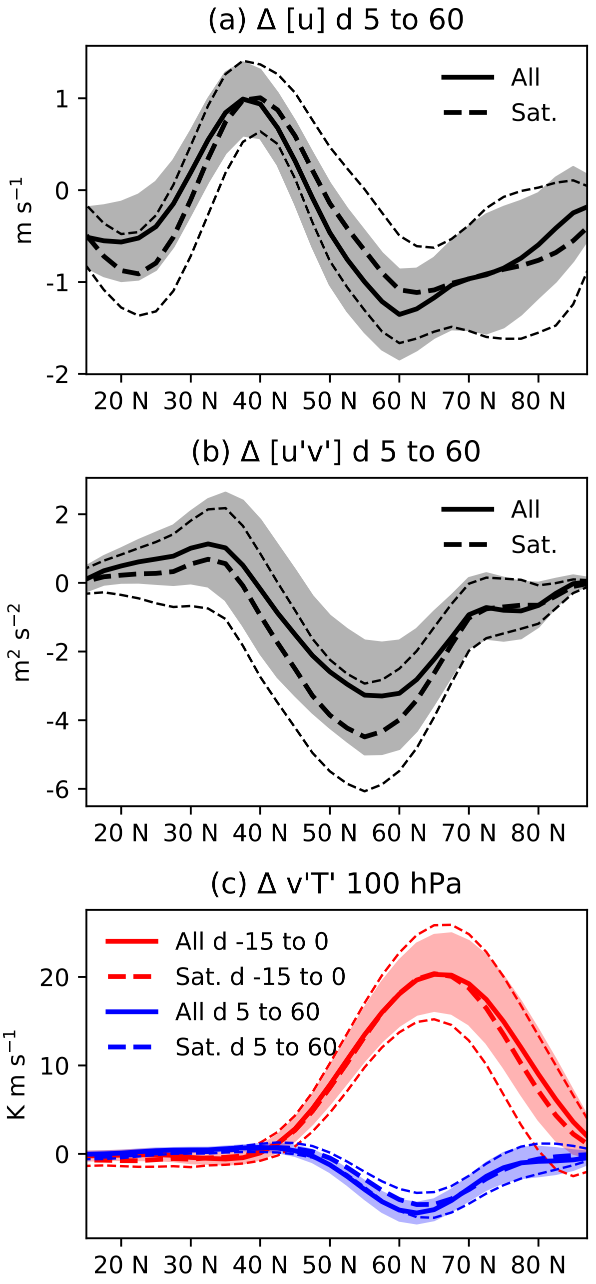

Figure 10. (a) Composite mean of vertically averaged zonal wind anomalies, averaged over lags 5 to 60 days following major warmings. Solid line shows the composite for all events while the dashed line shows the composite for the satellite era alone. Confidence intervals for the whole period are shaded while those for the satellite era are indicated by thin dashed lines. (b) Similar but for vertically integrated momentum fluxes. (c) Similar but for meridional heat fluxes at 100 hPa, averaged over lags -15 to 0 (in red), and over lags 5 to 60 (in blue). See text for details.

Table 1. Reanalysis products and dates considered in the present work. See Fujiwara et al. (2017) for a much more thorough discussion of the observations assimilated into each product. Abbreviations for certain products used within the text are indicated within parentheses.

Product (Label)	Reference	Centre	Dates considered	Classes of data assimilated
JRA-25	(Onogi et al., 2007)	JMA	01-1979 to 12-2010	All
JRA-55	(Kobayashi et al., 2015)	JMA	01-1958 to 12-2010	All
JRA-55C	(Kobayashi et al., 2014)	JMA	01-1979 to 12-2010	Conventional
MERRA	(Rienecker et al., 2011)	NASA GMAO	01-1979 to 12-2010	All
MERRA-2	(Gelaro et al., 2017)	NASA GMAO	01-1981[†] to 12-2010	All
ERA-40	(Uppala et al., 2005)	ECMWF	01-1958 to 08-2002	All
ERA-Interim	(Dee et al., 2011)	ECMWF	01-1979 to 12-2010	All
ERA-20C	(Poli et al., 2013)	ECMWF	01-1979 to 12-2010	Surface
NCEP-NCAR R1 (NCEP-NCAR)	(Kalnay et al., 1996)	NOAA/NCEP and NCAR	01-1979 to 12-2010	All
NCEP-DOE R2 (NCEP-DOE)	(Kanamitsu et al., 2002)	NOAA/NCEP and DOE	01-1979 to 12-2010	All
CFSR	(Saha et al., 2010)	NOAA/NCEP	01-1979 to 12-2010	All
NOAA-CIRES 20CR v2c (20CR v2)	(Compo et al., 2011)	NOAA and CIRES	01-1979 to 12-2010	Surface

[†] Although MERRA-2 includes 1980, there are spin-up issues in early 1980 which affect the Arctic vortex.