Referee 2

We are thankful to the reviewer for these comments which help us improve the quality of our manuscript. Our response to each comment is written below. We note that we have noticed a small coding error: averages were made between 50-85N instead of 45-85N as indicated in the manuscript. After correcting this mistake, there is a small change in the classification of High-Agreement and Low-Agreement events (please see the correction to the table where SSW events are listed) but this does not alter our interpretation of the results.

General comments

This is an interesting study comparing several reanalysis datasets to analyze the momentum equation during sudden stratospheric warmings (SSW). In particular, I find useful the idea of analyzing the latest reanalysis ensemble separately. Having so many available datasets, an evaluation of how they perform depending on the topic of the study is required. I have enjoyed reading the manuscript and I appreciate the advises and recommendations regarding the uses of different reanalysis in stratospheric dynamics. The contain of this manuscript would be valuable for the scientific community so I recommend it for publication. I only have some minor questions listed next.

Specific comments

Page 6, line 10: What happens if a SSW is detected in less than 4 datasets? Is it excluded? Does this

happen in many cases?

If the event is not detected in at least 4 datasets it is excluded, which we now indicate in the manuscript. It happens in February 81, February 95 and February 2002. This means only three events out of a total of 25.

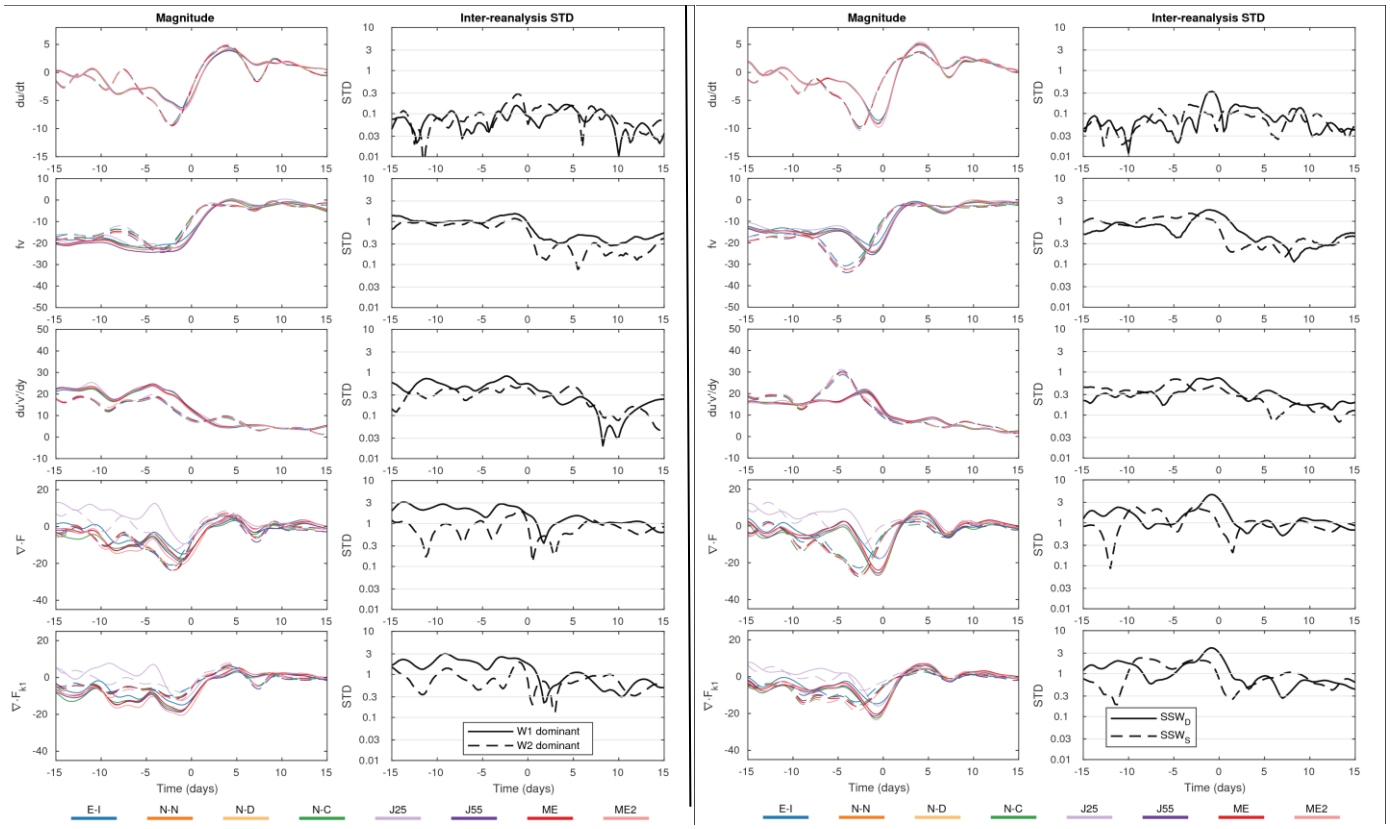
Table 2: Why the analysis stops in 2010? There is no data for some reanalysis considered herein?

The last SSW listed is in 2010 because we use the official comparison period of the S-RIP project which ranges from 1980 to 2012 (mentioned in the data section). No events are detected in 2011 or 2012. Only one more event would be included if the analysis is extended to 2013 after which JRA-25 is not provided.

Page 7, line 5 and Page 24, last paragraph: The authors argue that split and displacement SSW-types result from different planetary-scale wave forcing which motivates that distinction. I wonder whether it would be more meaningful to distinguish between wavenumber 1 and 2 as SSW precursors instead of the split/displacement sub-classification. As shown in previous studies (e.g., Bancala et al. JGR-2012), he ratio W2/W1 events will be much smaller than the split/displacement type, but it could be worthy to check whether differences are significant in that case.

Thank you very much for the comment. It is true that the geometry of the vortex (split or displacement) does not necessarily reflect the wave drag that produced the SSW event. To investigate the role of the longitudinal scale of wave activity fluxes to the spread among reanalysis, we classify events according to whether they are dominated by wavenumber-1 (W1) EP flux or wavenumber-2 (W2) EP flux. We indicate the outcome of this classification in Table 2 of the revised manuscript.

We have added the following paragraph in the manuscript: *On the other hand, when considering the dominant fluxes of wave activity producing SSW evens, we find that out of 7 LASSWs, four are W1-dominant and 1 is W2 dominant and out of 7 HASSWs, 1 is W1-dominant and 3 are W2-dominant. This seems to indicate that wavenumber-1 wave drag is responsible for larger uncertainties in reanalysis datasets but a detailed analysis reveals that inter-reanalysis spread is not markedly different between W1-dominant and W2-dominant events (not shown) suggesting that it is the intensity of wave drag rather than the longitudinal scale of wave activity that is linked to uncertainties among reanalyses.* Supplementary Fig. 1 shown here supports this analysis.



Supplementary Figure 1: Similar to Fig. 11 of the manuscript but for comparing SSWs that are primarily forced by wave-1 or wave-2 EP flux (solid and dashed, respectively; shown on left side) and SSWs that are displacements or splits (solid and dashed, respectively; shown on right side). Although there are small differences in the deceleration of zonal-mean zonal wind, the weakest stratospheric winds following the SSWs are similar in all types of events (not shown).

Page 12, line 8: What do you mean with "our interpretation of the evolution of SSW events"?

What we mean is that the uncertainties in the upper-stratosphere have a greater impact on how we understand the evolution of SSW events from observations. We rephrase this section to improve the clarity.

Technical corrections

Figure 2: It is difficult to distinguish the thin and thick lines. Maybe using dashed and solid lines as in figure 5?

We are now using dashed lines.

Figures 3 and 4: I have found difficult to distinguish contours with this color scale. Especially in a printed version.

We use a new color scale that makes interpreting these figures easier.

Figure 6 (right column): Note that some plots are out of range.

We extend the range of this figure.

Page 8, line 9: events or event?

Corrected

Page 10, line 25: quantities.,

Corrected

Page 15, 4: Here, and in other parts of the text: m/s/day. This notation is confusing, I would suggest using ms-1day-1

We now use $ms^{-1}day^{-1}$ throughout the text.

Referee 3

We are thankful to the reviewer for these comments which help us improve the quality of our manuscript. Our response to each comment is written below. We note that we have noticed a small coding error: averages were made between 50-85N instead of 45-85N as indicated in the manuscript. After correcting this mistake, there is a small change in the classification of High-Agreement and Low-Agreement events (please see the correction to the table where SSW events are listed) but this does not alter our interpretation of the results.

General Comments: This paper examines and compares the momentum budget during sudden stratospheric warming (SSW) events using eight reanalysis data sets. Their results provide some insights into the uncertainties of the budget equation during SSWs, especially the contributions of the QG and non-QG terms, the spread or the discrepancies in terms of the regions and the periods. It is also very useful to know that the spread is much reduced in the latest reanalysis products.

The authors suggested that the largest discrepancy originated mainly from the Coriolis torque (in abstract, line 13, page 7 and section 5). Momentum flux convergence is mentioned as the second term which presents non-negligible spread. I am concerned the word “originated”. This gives an impression that if we fix fv, we would get SSW right. However, the origin of the uncertainties must be in the wave forcing rather in the zonal mean meridional velocity, given the meridional circulation is driven primarily by wave forcing.

We agree with the reviewer that the choice of the word "originated" is not appropriate. The discrepancies in the Coriolis force may be due to discrepancies in wave drag (resolved and not resolved) but also due to biases in the mean state and data assimilation procedure (Kobayashi and Iwasaki, 2016; Uppala et al., 2005). We rephrase our statement to: While the largest uncertainties in the momentum budget are found in the Coriolis torque, momentum flux convergence also presents a non-negligible spread among the reanalyses.

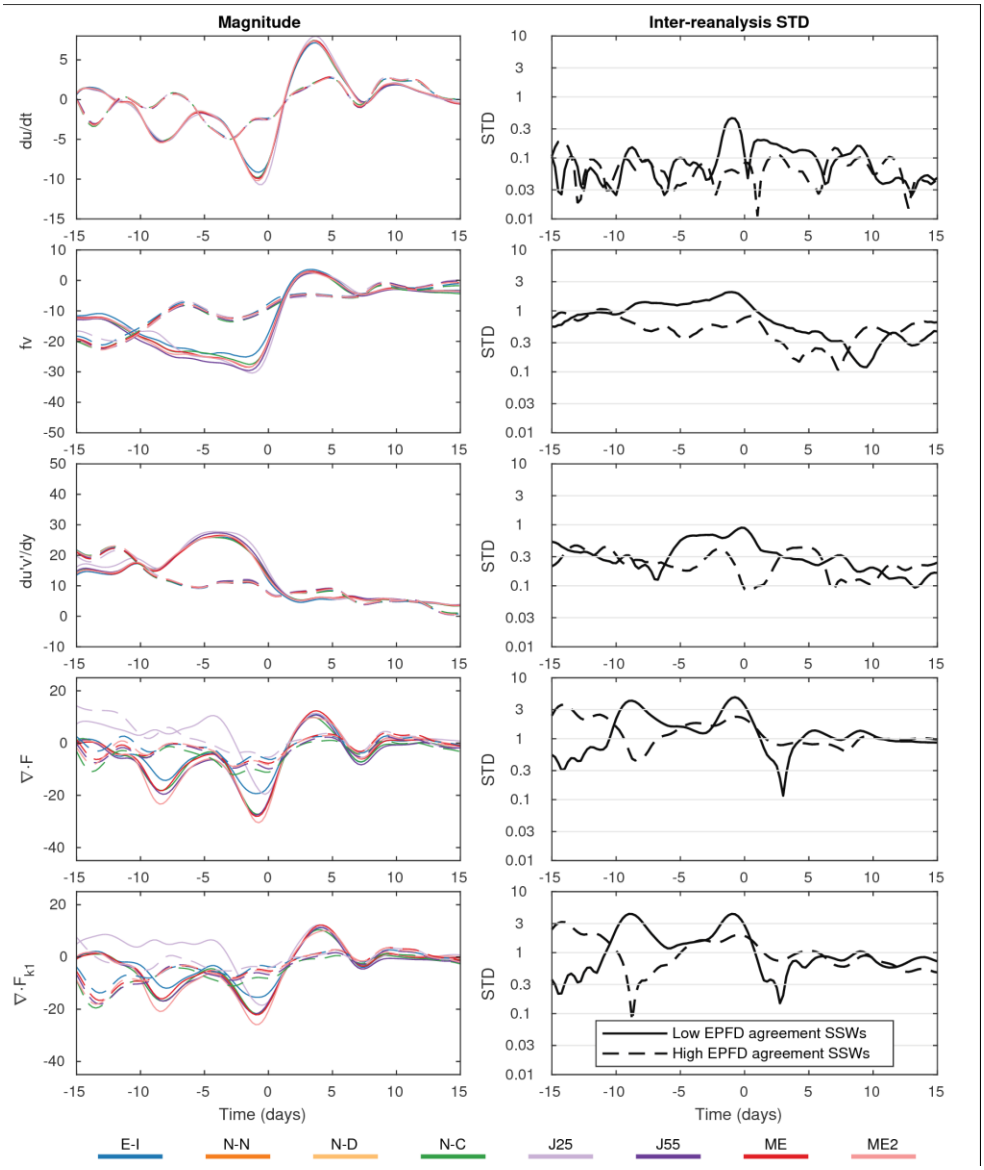
Also, what their results actually suggest is that the largest discrepancy is associated with the residual term R, the last term in equation (1.1). This can be seen clearly in their figures 5-7. The standard deviation associated with R is slightly smaller but comparable in magnitude to that of fv. However, the mean state of fv is one magnitude larger than R. Thus, R rather than fv has the largest discrepancies. I suggest that the authors make this point clearer by simply stating that the resolved part of the discrepancies is mainly associated with fv.

We agree that it should be clarified that the largest resolved discrepancy is associated with fv but that unresolved discrepancies, found in R, can also be large. We modify the abstract and conclusion accordingly.

I am concerned with their definition of high-agreement and low-agreement SSW events. Those events were defined by the standard deviation of the Coriolis torque averaged from 45-85N. fv is not even an effective measure for SSW events. There are times (within ~5-15 day average window) when zonal mean fv is large but there is no SSW. I do not think that it is appropriate to define the strength of a SSW event (i.e the strongest or weakest) or the associated discrepancies just by using fv. Again, this is because both SSW and the changes in fv and their associated uncertainties are consequences of wave mean-flow interaction.

Although the Coriolis torque produced by the meridional circulation is not a common measure for SSW events, our analysis is motivated by the fact that it is a major source of uncertainty of the momentum budget, and one of the major circulation changes that lead to SSW events. We did not state anywhere that we judged the strength of SSW events using fv. We noted that events with largest discrepancies in fv had a more intense deceleration of zonal-mean zonal wind. As the reviewer suggests, discrepancies in fv can result from discrepancies in wave drag, but also from biases in the mean state and data assimilation procedure (Kobayashi and Iwasaki, 2016; Uppala et al., 2005).

We show in supplementary Fig. 1 the outcome of classifying high agreement SSWs and low agreement SSWs using EP flux divergence, a measure of resolved wave drag, instead of the Coriolis torque. Similar to classifying events with the Coriolis torque, events that have larger discrepancies in wave drag show a stronger deceleration and stronger forcings by the Coriolis torque and momentum flux convergence, as well as stronger forcing for deceleration by EPFD. Since our focus was not on the terms of the transformed Eulerian mean momentum equation, we do not discuss of this analysis in the manuscript.



Supplementary Figure 1: Similar to Fig. 11 of the manuscript but for comparing SSWs with small (HASSWs – dashed lines) and large uncertainties (LASSWs – solid lines) based on EPFD instead of the Coriolis torque.

Other than the above points, the paper is well written in general. I suggest publication with some effort to improve the clarity of the expressions. More specific comments are provided below.

Specific comments:

1) Line 15, page 1: "the onset of SSW events, a period characterized by unusually large fluxes of planetary-scale waves from the troposphere to the stratosphere". This sentence holds true only if the period is ~40

days (Polvani and Waugh 2004). The correction between the wave fluxes (or v'T') would become much reduced if the averaging period is only 5-15 days, which is used in this study (i.e. figures 7-8 and figures 11-12). At these shorter time scales, stratospheric internal variation becomes important. This is precisely why the models cannot predict the timing or the initialization of SSWs. The authors must be careful when they discuss their results and when they related to the EP flux divergence to those from the troposphere.

Whereas Polvani and Waugh, (2004) noted that wave fluxes integrated over a period of 40 days were well correlated with the strength of the stratospheric polar vortex, other studies reported a strong link between short-lived bursts of planetary-scale wave activity and the rapid deceleration of the stratospheric polar vortex (Martineau and Son, 2015; McDaniel and Black, 2005; Sjoberg and Birner, 2014). We agree with the reviewer that the intrinsic variability of the stratosphere is important as it can significantly influence the amount of wave activity that can propagate from the troposphere but without a source of wave activity within the troposphere, vortex vacillations would not occur. Since the sentence the reviewer is referring to is supported by our results and previous studies, we decide to keep it as is in the revised manuscript.

2) Line 20, page 1: "The strongest SSWs being subject to larger discrepancies among reanalyses". This sentence gives one impression that there is an accepted definition of "the strongest SSWs". Naturally, the readers would think that these events produced the warmest temperature or strongest easterly winds. Is this true?

We agree that there is no commonly accepted definition of what is a strong SSW. We thus clarify that SSWs with the most intense deceleration of zonal-mean zonal wind show larger discrepancies among reanalysis data sets.

3) Line 15-16, page 3. It is better to state that the previous assessment was mainly for the extratropics. In the tropics where the QBO becomes important, higher vertical and horizontal resolution should lead to much improved dynamical consistency.

We clarify that this assessment was done in the extratropics. We agree that repeating the analysis in the tropical region may lead to different conclusions.

4) Line 21-22, page 4. The last term R also accounts for non-conservative processes, such as Rossby wave breaking (RWB). During SSW, planetary-scale RWB can play an important role. Interestingly, the largest error is associated with R rather than fv term.

The process of planetary-scale wave breaking (not small-scale gravity wave breaking) is largely conservative and resolved by reanalyses until wave activity is transferred to physical scales near the limit of what the model can resolve. Numerical diffusion, that we already mentioned in the manuscript, then dissipates wave activity at these small scales. This diffusion is included in R.

5) Lines 12-16, page 7. Now I understand that the definition is based on the largest discrepancies in the Coriolis torque. This needs to be made clearer in the abstract when you mentioned the strongest SSWs because there is no such a definition in terms of the known or accepted description of the SSWs. Also, see my general comments for further concerns.

We now clarify that the uncertainties in the Coriolis torque are larger when SSW events display a more intense deceleration of the stratospheric polar vortex.

6) Line 2, page 8. "The evolution of geopotential height contours". Please include the values here (not just in the figure caption) and justify why those values are used to describe the polar vortex. Ertel Potential vorticity should be a much better quantity for this purpose and why not to use EPV?

We now mention the values of these contours in the text. The reason we chose these contours is simply because they clearly illustrate the shape of the stratospheric polar vortex throughout the life cycle of the 2009 SSW event and we now mention it in the manuscript. We agree that EPV is a better quantity to study the dynamical evolution of SSWs as it is conserved for conservative flows. However, we wished to use a less derived quantity to illustrate the shape of the stratospheric polar vortex. Geopotential height serves our purpose well as it is parallel to the geostrophic flow. Geopotential height is used frequently to describe the evolution of SSW events (Charlton and Polvani, 2007; Limpasuvan et al., 2004; Martineau and Son, 2015; Seviour et al., 2013).

7) Figures 3-4. It is really hard to qualify the spread or discrepancies based on the color bar used.

We are now using a better-suited colormap.

8) Line 7, page 15. "Terms that are left of the QG from of the momentum equation provide much smaller forcing for zonal wind tendency during SSW events . . . Their differences from one reanalysis ". I disagree for the following reasons. 1). The two QBO terms are of the opposite sign in general (see figure 7). If they are added together, the sum would have a comparable magnitude when it is compared with the other terms. 2). It is well-known that the SSW events often involve breaking of finite amplitude waves.

Such an effect cannot be accounted for by 2.5 resolution pressure level data. Please reword the part to avoid the possibility of misleading the readers. See my general comments for further information.

Even when the two QG terms, which are often opposed, are added together, their total contribution is usually larger than the non-QG terms and explain the largest fraction of zonal wind tendencies (see supplementary Fig. 3). But, we agree with you that it may be more accurate to tone down our affirmation by replacing *much smaller* by *smaller*.

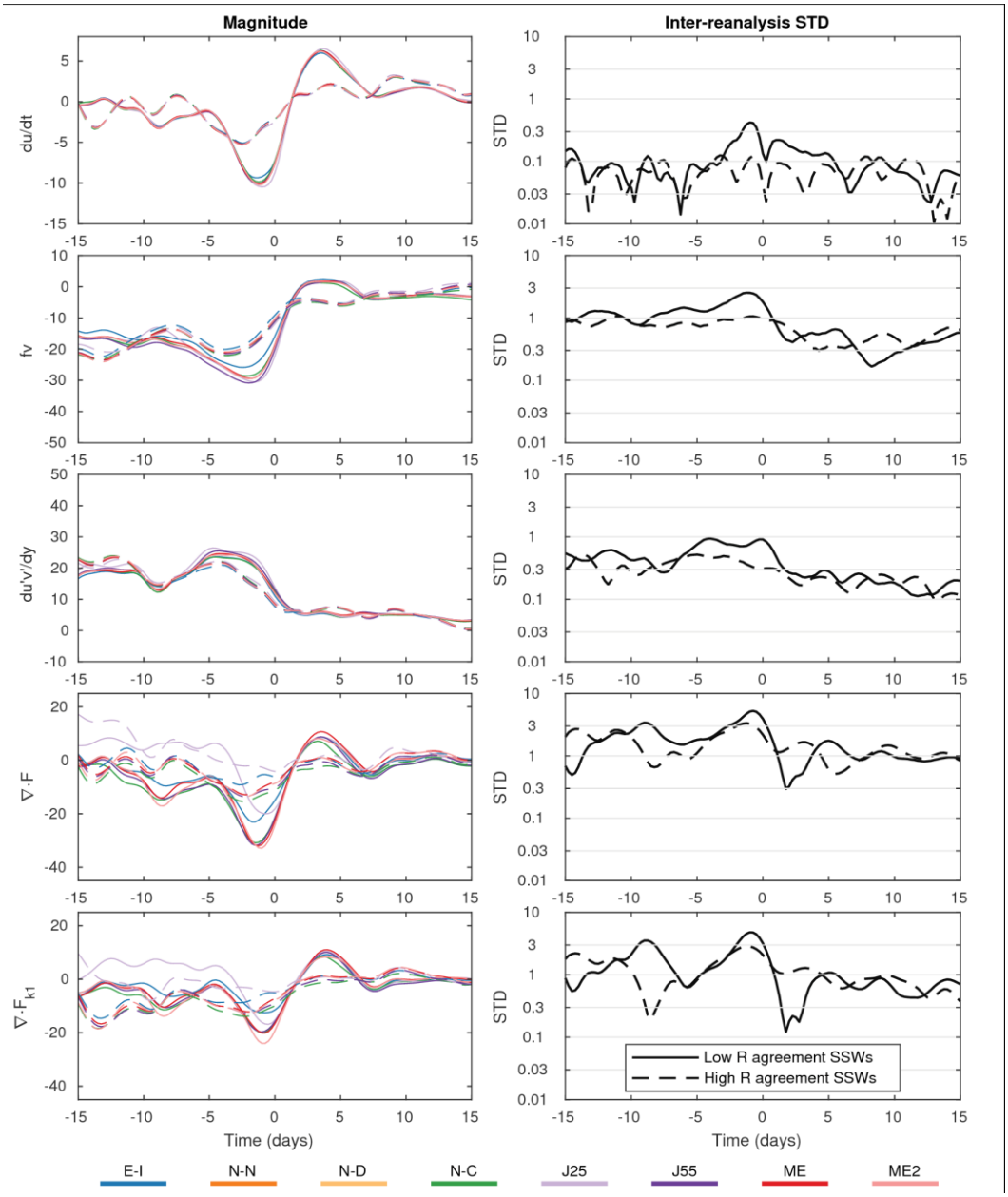
Most of the deceleration during SSW events results from a large amplification of planetary-scale waves in the stratosphere (Martineau and Son, 2015; Solomon, 2014). The deceleration of zonal wind is strongest when the QGPV field is deformed by wave-1 or wave-2 disturbances (v'q'). When the waves reach large amplitudes, wave-breaking will indeed redistribute wave activity to smaller scales (this does not necessarily lead to meridional QGPV fluxes and deceleration of zonal-mean zonal wind). As mentioned earlier, once wave activity is transferred to smaller scales, it will be dissipated by numerical diffusion and be included in the residual term. Overall, most of the deceleration of zonal wind during SSW event is well accounted for by the QG terms in reanalyses (Martineau and Son, 2015).

9) I am not sure whether or not figures 8 and 9 is needed. Would it be more concise or informative if the figures were combined as one and show the two groups: the latest versus older generation reanalysis products?

We decide to keep these figures since they may be useful for reanalysis centers and reanalysis users to evaluate discrepancies of specific reanalyses with respect to others and locate the regions of the atmosphere responsible for these biases. This would not be possible by showing only composites of newer versus older reanalyses.

10) Lines 18-28, Page 22. I suggest that the authors to check would the same spread or results be obtained using the residual term R and its standard deviation to define HASSWs and LASSWs. Same applies to figures 11 and 12.

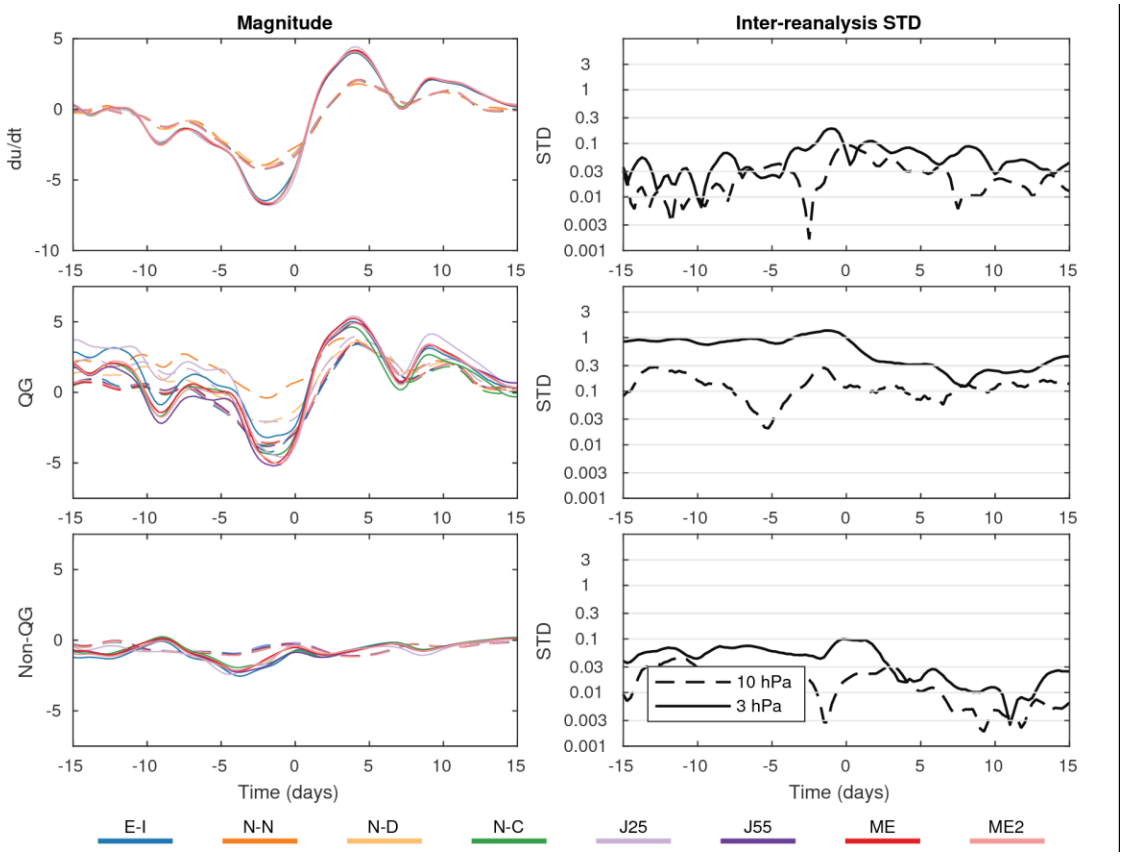
As suggested by the reviewer, we verify if the same result would hold by defining HASSWs and LASSWs using the residual of the momentum budget instead of the Coriolis torque. The results are shown here with supplementary Fig. 2. Events with large discrepancies of the residual show a more intense deceleration, although the forcings by the Coriolis torque and momentum fluxes are more similar compared to the differences between HASSWs and LASSWs defined with the Coriolis torque. This suggests that what differentiates these two categories of events is most likely the strength of unresolved forcing included in R, such as gravity wave drag. Since unresolved forcings are not the foci of this work, we elect to not discuss of this result further in the manuscript.



Supplementary Figure 2: Similar to Fig. 11 of the manuscript but for comparing SSWs with small (HASSWs – dashed lines) and large uncertainties (LASSWs – solid lines) based on R instead of the Coriolis torque.

11) Line 14, page 26. See general comments. The results do not suggest that the discrepancies in those non-QG terms are smaller than the QG-terms.

We show in supplementary Fig. 3 that the sum of QG terms is typically larger than the sum of non-QG terms. We agree that the difference may not be as large as we suggest and thus will tone down our statements wherever applicable.



Supplementary Figure 3: Same as Fig. 5 of the manuscript except that the QG and non-QG terms are summed together.

12) Line 23, page 26. "Most of the residual in the stratosphere is correlated to uncertainties in the Coriolis torque". This is very interesting and somehow expected. My explanation is as follows. In the upper stratosphere, gravity wave breaking and finite amplitude wave activities appear regularly there but their propagation cannot be well captured by 2.5 degree pressure level data. Their effects on the polar vortex or zonal mean zonal wind would be included in R or the vertical momentum flux term especially when the QG-terms are calculated by using variables such as u and v, as it is done by this study. On the other hand, when the EP flux divergence is included as in the transformed Eulerian mean equations, the variation of wave forcing would be better resolved by the data used. This is because Del F accounts for the vertically propagating wave not just the meridionally propagating waves. This is confirmed by figure 11. The figures shows, at 3 hPa, the temporal evolution of the zonal mean wind tendency follows better with the EP flux

divergence, less so in terms of fv. Thus, I would think that it is the uncertainties associated with non-resolved wave forcing caused the spread in fv, rather than the other way around.

The bulk of the deceleration of zonal-mean zonal wind in the stratosphere is produced by planetary-scale waves and is thus well resolved by 2.5 degree pressure level data (most of meridional fluxes of QGPV, and thus EP flux convergence, are resulting from large scale motion by wave-1 and wave-2 planetary-scale waves). We agree however that wave-breaking may lead to fluxes at scales smaller than what can be resolved by a 2.5 degree grid and that the effect of gravity waves, which is not included in our diagnostic, will show up in R.

We agree that the spread in fv may result from discrepancies in wave drag, either from planetary waves or gravity waves and will make sure to clarify this point in the revised manuscript. Reanalysis datasets, however, are not only following their own physics, they assimilate data and are also susceptible to discrepancies in data assimilation (Kobayashi and Iwasaki, 2016; Kobayashi et al., 2015; Uppala et al., 2005). We make modifications to the conclusion to take into account these contributions.

Minor comments:

1) Line 29, page 1. Too many citations here for motivation.

We reduce the number of citations.

2)Line 5, page 2. Two daughter vortices -> two vortices.

Corrected

3) Line 6, page 2. Please be more specific about the differences. Otherwise, delete the sentence as it adds no information.

We remove this sentence.

4) Line 9, page 2. “the general signature”. What is it? Please be more

specific.

We now clarify what is this signature by mentioning the deceleration of zonal-mean zonal wind and the warming of the polar cap.

5) Line 25, ., -> ,

Corrected

6) Line 16, high stratosphere -> upper stratosphere.

Corrected

Charlton, A. J. and Polvani, L. M.: A New Look at Stratospheric Sudden Warmings. Part I: Climatology and Modeling Benchmarks, J. Clim., 20(3), 449–469, doi:10.1175/JCLI3996.1, 2007.

Kobayashi, C. and Iwasaki, T.: Brewer-Dobson circulation diagnosed from JRA-55, J. Geophys. Res. Atmos., 121(4), 1493–1510, doi:10.1002/2015JD023476, 2016.

Kobayashi, S., Ota, Y., Harada, Y., Ebita, A., Moriya, M., Onoda, H., Onogi, K., Kamahori, H., Kobayashi, C., Endo, H., Miyaoka, K. and Takahashi, K.: The JRA-55 Reanalysis: General Specifications and Basic Characteristics, J. Meteorol. Soc. Japan. Ser. II, 93(1), 5–48, doi:10.2151/jmsj.2015-001, 2015.

Limpasuvan, V., Thompson, D. W. J. and Hartmann, D. L.: The Life Cycle of the Northern Hemisphere Sudden Stratospheric Warmings, J. Clim., 17(13), 2584–2596, doi:10.1175/1520-0442(2004)017<2584:TLCOTN>2.0.CO;2, 2004.

Martineau, P. and Son, S.-W.: Onset of Circulation Anomalies during Stratospheric Vortex Weakening Events: The Role of Planetary-Scale Waves, J. Clim., 28(18), 7347–7370, doi:10.1175/JCLI-D-14-00478.1, 2015.

McDaniel, B. A. and Black, R. X.: Intraseasonal Dynamical Evolution of the Northern Annular Mode, J. Clim., 18(18), 3820–3839, doi:10.1175/JCLI3467.1, 2005.

Polvani, L. M. and Waugh, D. W.: Upward Wave Activity Flux as a Precursor to Extreme Stratospheric Events and Subsequent Anomalous Surface Weather Regimes, J. Clim., 17(18), 3548–3554, doi:10.1175/1520-0442(2004)017<3548:UWAFAA>2.0.CO;2, 2004.

Seviour, W. J. M., Mitchell, D. M. and Gray, L. J.: A practical method to identify displaced and split stratospheric polar vortex events, Geophys. Res. Lett., 40(2), 5268–5273, doi:10.1002/grl.50927, 2013.

Sjoberg, J. P. and Birner, T.: Stratospheric wave-mean flow feedbacks and sudden stratospheric warmings in a simple model forced by upward wave-activity flux, J. Atmos. Sci., (1976), 141006071034003, doi:10.1175/JAS-D-14-0113.1, 2014.

Solomon, A.: Wave Activity Events and the Variability of the Stratospheric Polar Vortex, J. Clim., 27(20), 7796–7806, doi:10.1175/JCLI-D-13-00756.1, 2014.

Uppala, S. M., Kallberg, 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. M., Berg, L. Van De, Bidlot, J., Bormann, N., Caires, S., Chevallier, F., Dethof, A., Dragosavac, M., Fisher, M., Fuentes, M., Hagemann, S., Holm, E., Hoskins, B. J., Isaken, 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 re-analysis, Q. J. R. Meteorol. Soc., 131, 2961–3012, 2005.

A comparison of the momentum budget in reanalysis datasets during sudden stratospheric warming events

Patrick Martineau[1], Seok-Woo Son[2], Masakazu Taguchi[3], Amy H. Butler[4]

[1]Research Center for Advanced Science and Technology, The University of Tokyo, Tokyo, Japan
[2]School of Earth and Environmental Sciences, Seoul National University, Seoul, South Korea
[3]Department of Earth Science, Aichi University of Education, Kariya, Japan
[4]University of Colorado Cooperative Institute for Research in Environmental Sciences

Correspondence to: Patrick Martineau (pmartineau@atmos.rcast.u-tokyo.ac.jp)

Abstract. The agreement between reanalysis datasets, in terms of the zonal-mean momentum budget, is evaluated during sudden stratospheric warming (SSW) events. It is revealed that there is a good agreement among datasets in the lower stratosphere and troposphere concerning zonal-mean zonal wind, but less so in the upper stratosphere. Forcing terms of the momentum equation are also relatively similar in the lower atmosphere, but their uncertainties are typically larger than uncertainties of the zonal wind tendency. Similar to zonal wind tendency, the agreement among forcing terms is degraded in the upper stratosphere. Discrepancies among reanalyses increase during the onset of SSW events, a period characterized by unusually large fluxes of planetary-scale waves from the troposphere to the stratosphere, and decrease substantially after the onset. While the largest uncertainties in the resolved terms of the momentum budget ~~originate from~~are found in the Coriolis torque, momentum flux convergence also presents a non-negligible spread among the reanalyses. Such a spread is reduced in the latest reanalysis products, decreasing the uncertainty of the momentum budget. It is also found that the uncertainties ~~of the momentum budget~~in the Coriolis torque depend on the strength of SSW events: the ~~strongest~~ SSWs that exhibit the most intense deceleration of zonal-mean zonal wind~~being~~ are subject to larger discrepancies among reanalyses. These uncertainties in stratospheric circulation, however, are not communicated to the troposphere.

1 Introduction

Sudden Stratospheric Warming (SSW) events are prime manifestations of coupling between the tropospheric and stratospheric circulations (Baldwin, 2001). They are characterized by a rapid deceleration and reversal of the stratospheric zonal wind resulting from an enhanced injection of planetary-scale wave activity from the troposphere to the stratosphere (Limpasuvan et al., 2004; Martineau and Son, 2015; Polvani and Waugh, 2004). The changes in stratospheric circulation can then, in return, influence tropospheric weather (Kidston et al., 2015). Motivated by their role for tropospheric predictability (Sigmond et al., 2013; Taguchi, 2015; Tripathi et al., 2015), SSWs have been the object of many observational and modelling studies.

Recent studies have highlighted a sensitivity to the choice of reanalysis dataset for the detection of SSW events. For instance, Charlton and Polvani (2007) noted discrepancies in the central date of SSW events

between ERA-40 and NCEP-NCAR. They also found discrepancies in the classification of those events per the geometry of the distorted stratospheric polar vortex, whether it is displaced off the pole or split into two ~~daughter~~ vortices. ~~Song and Chun (2016) also found some differences in ERA-Interim, JRA-55, and MERRA.~~ Other definitions of SSW events also show discrepancies among reanalyses (Butler et al., 2015). Since SSWs are often defined using a threshold value, requiring a reversal of zonal-mean zonal wind ($\overline{u} = 0$) at 10 hPa and 60°N (Charlton and Polvani, 2007), their detection can be sensitive to small differences in the zonal wind between reanalyses. Despite those discrepancies, Palmeiro et al. (2015) have noted that the ~~general signature~~main features of SSW events, such as the deceleration of zonal-mean zonal wind and warming of the polar cap, ~~is~~are not sensitive to the choice of reanalysis. Furthermore, composites of the northern annular mode (NAM) index ~~during~~ for a common set of SSW event dates ~~for all reanalyses~~ were shown to be similar (Martineau and Son, 2010).

Despite the seemingly good agreement of zonal-mean zonal-wind, temperature, and geopotential height between datasets during SSW events, inter-dataset variability merits further investigation. Lu et al. (2015) recently highlighted non-negligible differences in the wave drag and the residual circulation between ERA-40 and ERA-Interim. A comprehensive comparison of momentum diagnostics in the stratosphere among reanalyses further revealed non-negligible inconsistencies in the zonal-mean momentum equation, resulting primarily from inter-data variability in the residual circulation in the mid-stratosphere (Martineau et al., 2016). Such variability of the Brewer-Dobson circulation among reanalysis datasets is well documented in the literature (Abalos et al., 2015; Iwasaki et al., 2009; Monge-Sanz et al., 2013).

Martineau et al. (2016) have also shown that the ability to explain the stratospheric zonal-mean zonal wind tendencies using the forcing terms of the zonal-mean momentum equation has improved in the latest reanalysis products and that momentum diagnostics tend to agree better among the latest reanalyses. This improvement was demonstrated in the context of the wintertime climatology and different regimes of vortex variability (strong/weak, accelerating/decelerating) with an emphasis on the mid-stratosphere. In this work, we focus on the most extreme events of stratospheric variability, the SSW events, and extend the comparison of the momentum budget to the upper stratosphere and troposphere.

The reanalysis datasets evaluated in this work are first presented in Section 2.1, followed by a description of the momentum diagnostics used throughout this study in Section 2.2. The SSW events used for the comparison of reanalyses are then presented in Section 2.3. The uncertainty of stratospheric vortex structures among reanalysis datasets is first evaluated in Section 3 using as a case study the January 2009 SSW event. Then, zonal-mean quantities and diagnostics using the zonal-mean momentum equation are shown in Section 4. Conclusions are finally presented in Section 5.

2 Methodology

2.1 Data

The eight reanalysis datasets compared in this study are listed in Table 1. ERA-40 from the European Centre for Medium-Range Weather Forecasts (ECMWF) is excluded as it is not provided for recent years, thus limiting the sample size of SSW events since 1980. The NOAA Twentieth Century Reanalysis (20CR) and the ECMWF Twentieth-Century Reanalysis (ERA-20C) are also left out as they are known to have unrealistic stratospheric variability in comparison to reanalyses that fully assimilate upper atmospheric observations (Compo et al., 2011; Poli et al., 2013). Temperature and the three-dimensional wind field are used on pressure levels for each reanalysis. To prevent our diagnostics from being affected by the vertical resolution, which gives an unfair advantage to the latest reanalyses, only 22 common vertical levels are kept (i.e., 1000, 925, 850, 700, 600, 500, 400, 300, 250, 200, 150, 100, 70, 50, 30, 20, 10, 7, 5, 3, 2, and 1 hPa) except for NCEP-NCAR and NCEP-DOE which are available only up to 10 hPa. Similarly, to ensure that the diagnostics are not affected by differences in horizontal resolution, each dataset is interpolated onto a standardized 2.5° by 2.5° grid, which is the coarsest grid provided among all the reanalyses considered in this study. This reduction of the resolution in some reanalyses is not expected to have a large impact on our comparison. In fact, in the extratropics, horizontal and vertical resolutions were previously shown not to have a substantial effect on momentum diagnostics except near the tropopause and in the upper stratosphere where higher vertical and horizontal resolutions improved slightly the dynamical consistency (Martineau et al., 2016). The reanalysis datasets are compared for a common period ranging from 1980 to 2012, 1980 being the first year for which MERRA2 data is provided and 2012 being the official final year for the comparison of reanalyses in the SPARC Reanalysis Intercomparison Project (S-RIP) project. An introduction to S-RIP and a comprehensive description of the reanalyses are provided in Fujiwara et al. (2017).

Table 1: Summary of reanalysis datasets included in the comparison

NAME	Label	Highest level (hPa)	Original Resolution[$]	Reference
ERA-Interim[LRE]	E-I	1	1.5	Dee et al. (2011)
NCEP–NCAR Reanalysis-1*	N-N	10	2.5	Kalnay et al. (1996)
NCEP–DOE Reanalysis-2	N-D	10	2.5	Kanamitsu et al. (2002)
NCEP Climate Forecast System Reanalysis[LRE,+]	N-C	1	2.5	Saha et al. (2010) and Saha et al. (2014)
JRA-25	J25	1	2.5	Onogi et al. (2007)
JRA-55[LRE]	J55	1	1.25	Kobayashi et al. (2015)
MERRA	ME	0.1	1.25	Rienecker et al. (2011)
MERRA2[LRE]	ME2	0.1	1.25	Gelaro et al. (2017)

* Vertical velocity is not provided above 100 hPa

[$] The original resolution is not necessarily the highest resolution provided by each reanalysis center.

[LRE] Included in the Latest Reanalysis Ensemble (LRE)

[+] Transition from version 1 to version 2 on January 1 2011

Whenever the average or standard deviation of multiple reanalyses is taken as a reference, it is performed on a subset including the latest reanalysis products from each center (ERA-Interim, NCEP-CFSR, JRA-55 and MERRA2). This composite is referred to as the latest reanalysis ensemble (LRE). In some figures, the LRE is contrasted to a composite of all reanalyses which we denote as the all reanalysis ensemble. The LRE subset emphasizes the discrepancies affecting the reanalysis datasets that are nowadays most commonly used in research while excluding older reanalyses whose deficiencies are well documented in the literature.

2.2 Momentum diagnostics

The zonal-mean momentum equation, derived from the primitive version of the equation in pressure coordinate, is expressed as

$$\frac{\partial \overline{u}}{\partial t} = f\overline{v} - \underbrace{\frac{1}{a\cos^2\phi}\frac{\partial(\cos^2\phi\,\overline{u'v'})}{\partial y}}_{fv} - \underbrace{\overline{v}\frac{1}{a\cos\phi}\frac{\partial(\overline{u}\cos\phi)}{\partial \phi}}_{Adv_\phi} - \underbrace{\overline{\omega}\frac{\partial \overline{u}}{\partial p}}_{Adv_p} - \underbrace{\frac{\partial(\overline{u'\omega'})}{\partial p}}_{dudw/dp} + R \quad (1.1)$$

where f is the Coriolis parameter, u, v, ω are the zonal, meridional, and vertical components of wind, ϕ is the latitude, and p is the pressure. Overbars and primes denote zonal mean and anomalies with respect to the zonal mean, respectively. While the left-hand side term expresses the zonal-mean zonal wind tendency, terms of the right-hand side represent forcing terms. They are, in order, the acceleration due to the Coriolis torque, the meridional convergence of momentum fluxes, the advection of zonal momentum by the meridional wind, the vertical advection of zonal momentum by the vertical wind, and the vertical convergence of vertical momentum fluxes. The last term, R, is referred to as the residual and represents sub-grid scale processes such as gravity wave drag and numerical diffusion. It also includes imbalances in the momentum equation introduced by the data assimilation process (analysis increment), errors due to the interpolation from model levels to pressure levels, and errors related to the numerical methods employed to evaluate each term of the equation. R can be used to quantify and compare the consistency of the momentum budget among reanalysis datasets (Lu et al., 2015; Martineau et al., 2016; Smith and Lyjak, 1985). The quasigeostrophic (QG) version of the momentum equation is often applied in the extratropics. Under this approximation, only the first two terms on the right-hand side of Eq. (1.1) are retained. The validity of this approximation can be assessed by comparing the magnitude of non-QG terms (3rd to 5th right hand side terms of Eq. (1.1)) to the QG terms. The numerical methods employed to evaluate Eq. (1.1) are described in further details in Martineau et al. (2016). Abbreviations of the various terms of Eq. (1.1) used in figures throughout this paper are indicated with braces in Eq. (1.1). In this work, the Eulerian-mean form of the momentum equation is preferred over the transformed Eulerian mean since additional vertical derivatives are needed in the latter, which can introduce numerical errors. The dynamical processes responsible for the eddy fluxes and the dominant terms of the momentum budget during SSW events are illustrated schematically in Fig. 1.

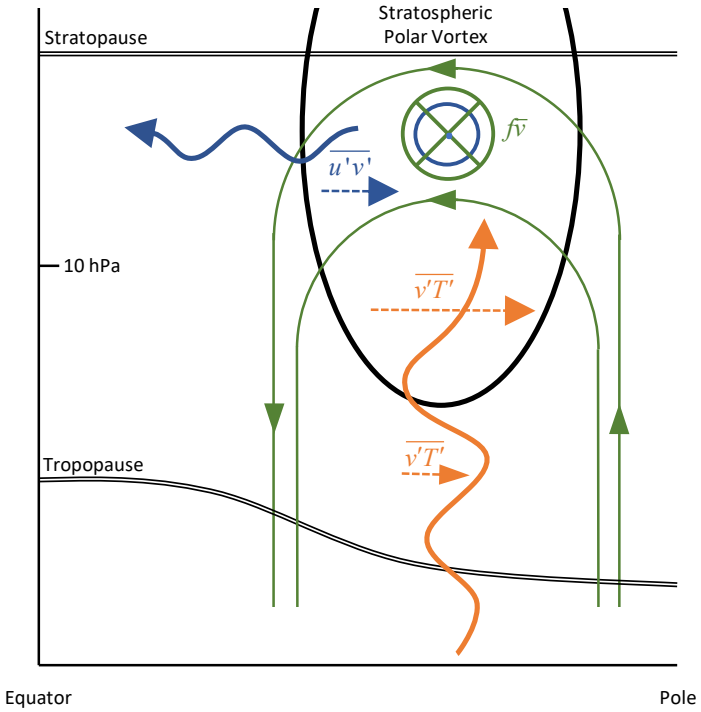


Figure 1: Schematic illustration of the dominant forcing terms of the momentum budget in the Eulerian framework during SSW events and their underlying dynamical processes. Planetary-scales waves propagate upward (orange) from the troposphere to the stratospheric polar vortex (thick black line) and are accompanied with poleward heat fluxes ($\overline{v'T'}$) amplifying with height. In agreement with the Matsuno (1971) model, the poleward heat fluxes generate a thermally direct circulation (green) with air rising at the pole and sinking in the mid-latitudes. To close this circulation, equatorward motion is generated in the stratosphere. The Coriolis torque resulting from this Circulation decelerates the wind (into the page symbol). Equatorward wave propagation is also observed (blue), accompanied with poleward fluxes of westerly momentum ($\overline{u'v'}$). The resulting convergence of momentum fluxes in the polar stratosphere acts to accelerate the zonal-mean zonal wind (out of the page symbol), counteracting partly the deceleration by the Coriolis torque. Note that in the transformed Eulerian mean the residual circulation is opposite, i.e., poleward in the stratosphere and downward at the pole ~~(Song and Chun, 2016)~~. The residual circulation approximates Lagrangian-mean motion.

2.3 Definition of SSW events

SSW events are defined following Charlton and Polvani (2007). Their central dates are set as the dates when the zonal wind reverses direction ($\overline{u} = 0$) in winter at 10 hPa and 60°N. Contrary to the original WMO definition, we do not verify if a reversal of the zonal-mean temperature gradient occurs at the same time as wind reversal. This additional criterion affects minimally the outcome of the detection (Charlton and Polvani, 2007). As mentioned earlier, since the definition of SSW events is based on a threshold value, the detection of a SSW can be sensitive to small variations in the wind field. Because of large variations in the characteristics of individual SSW events (e.g., Ayarzaguena et al., 2011; Martineau and Son, 2013), the comparison of reanalysis datasets could be negatively affected by using a different

set of events for each reanalysis. To ensure a fair comparison, SSWs are first detected in each reanalysis independently. If at least 4 reanalyses detect the SSW event, the onset date is set by averaging across the dates given by each reanalysis~~.~~. On a total of 25 events, three are detected in less than 4 reanalyses and are thus rejected. Detected event dates generally do not vary by more than 1-2 days between reanalyses. The common dates used for this comparison are listed in Table 2.

Table 2: Dates and types of SSW events used for the comparison. HA and LA denote high-agreement and low-agreement SSW events, respectively.

SSW onset date	Event type: split (S) or displacement (D)	Wavenumber forcing classification	Agreement
29-Feb-1980	D	W2	LA
04-Mar-1981	D		HA
04-Dec-1981		W1	~~HA~~
24-Feb-1984	D	W1	
01-Jan-1985	S	W2	
23-Jan-1987	D	W1	
08-Dec-1987	S		LA
14-Mar-1988	S	W2	HA
21-Feb-1989	S	W2	
15-Dec-1998	D	W1	LA
26-Feb-1999	S	W1	LA
20-Mar-2000		W2	
11-Feb-2001	S	W2	HA
31-Dec-2001	S	W1	LA
18-Jan-2003	S		HA
05-Jan-2004	D	W1	HA
21-Jan-2006	D		LA
24-Feb-2007	D	W2	HA
22-Feb-2008	D		LA
24-Jan-2009	S	W2	
09-Feb-2010		W1	
24-Mar-2010	D		HA
SSW (22)	D (10) S(9)	W1 (8) W2 (8)	LA (7) HA (7)

SSWs are also known to present a large diversity in terms of how the stratospheric polar vortex is distorted in the course of the events. While some events occur due to a displacement of the vortex, others occur from a splitting (Charlton and Polvani, 2007). These two types of SSW events result from different planetary-scale wave forcing in the stratosphere and can affect the tropospheric flow in different ways (Bancalá et al., 2012; Lehtonen and Karpechko, 2016; Martineau and Son, 2015; Mitchell et al., 2013; Smith and Kushner, 2012). It is thus possible that one type or the other is subject to larger uncertainties in reanalysis datasets. To test whether this is the case, the two types of SSWs, i.e., split SSWs (SSWS) and displacement SSWs (SSWD), are classified using vortex moment diagnostics described in Seviour et al. (2013) (see Table 2). ~~(~~Bancalá et al., (2012), however, pointed out that some

wavenumber-2 SSW events were not purely forced by fluxes of wavenumber-2 wave activity but that wavenumber-1 fluxes also contributed to the preconditioning of these events. We therefore also classify events according to the dominant forcing prior to the onset date. To this end, we compute the ratio of wavenumber-1 to wavenumber-2 vertical EP flux at 100 hPa over 14 days preceding the onset date. The eight events that have the largest ratio are classified as W1-dominant event and the eight events with the smallest ratio are classified as W2-dominant events. The outcome of this classification is indicated in Table 2.

SSWs are further classified based on whether the reanalysis datasets included in the LRE agree or disagree for a specific diagnostic. Since the Coriolis torque is the forcing term that shows the largest discrepancies among reanalysis datasets in the stratosphere (Martineau et al., 2016), it is used here to quantify the level of agreement. The standard deviation of the Coriolis torque averaged from 45°N to 85°N among the LRE is considered each day from 10 days before the central date to 5 days after the central date which is shown later to be the period when the Coriolis force is largest during SSW events. The agreement is then defined for each event as the maximum standard deviation observed during the period considered. Events are finally classified into two categories: high-agreement SSWs (HASSWs) or low-agreement SSWs (LASSWs) whether they are in the lower 33.3% or upper 33.3% of the agreement index of all SSWs. The type of each event is indicated in Table 2.

3 Vortex geometry during the 2009 SSW

Since the geometry of the stratospheric vortex may vary widely from one event to another, the structure of any particular event can be heavily obscured when performing composite means. We therefore first proceed to illustrate and compare the morphology of the stratospheric vortex among reanalysis datasets for a representative SSW event that is well documented in the literature. We choose the January 2009 SSW, an event characterized by a vortex splitting that resulted from an unusually large amount of upward EP flux by wavenumber two (Harada et al., 2010; Manney et al., 2009). Reanalysis datasets show qualitatively similar downward coupling during this event (Martineau and Son, 2010). Figure 2 illustrates the evolution of the 30 km and 38.5 km geopotential height contours at 10 and 3 hPa, respectively. These contours are chosen because they clearly illustrate the distortion of the stratospheric polar vortex during the life cycle of the 2009 SSW. In the course of this event, the polar vortex is progressively elongated to finally split into two individual vortices over northeastern Canada and Russia. The vortex structure features a westward tilt before the onset date (compare geopotential height contours at 10 and 3 hPa), which is consistent with the upward EP fluxes seen during the event (Harada et al., 2010).

The structure of the vortex is generally quite similar among reanalyses at 10 hPa, although some small discrepancies are observed in NCEP-NCAR and NCEP-DOE. Larger differences are found at 3 hPa. Notably, NCEP-CFSR exhibits contours enclosing a smaller area from two days before the onset (t_{-2}) and onward. On the other hand, JRA-25 exhibits contours enclosing a larger area, especially from t_{-2} to t_2.

From inspecting the 2009 SSW event~~s~~, it is clear that uncertainties in vortex geometry are larger in the upper stratosphere. These differences in geopotential height field are likely accompanied with differences in circulation, and thus, with differences in eddy activity and fluxes among the reanalyses. Although it is not an easy task to summarize differences in vortex geometry for all SSW events, it is possible to evaluate them indirectly through eddy fluxes, which are examined in the following section.

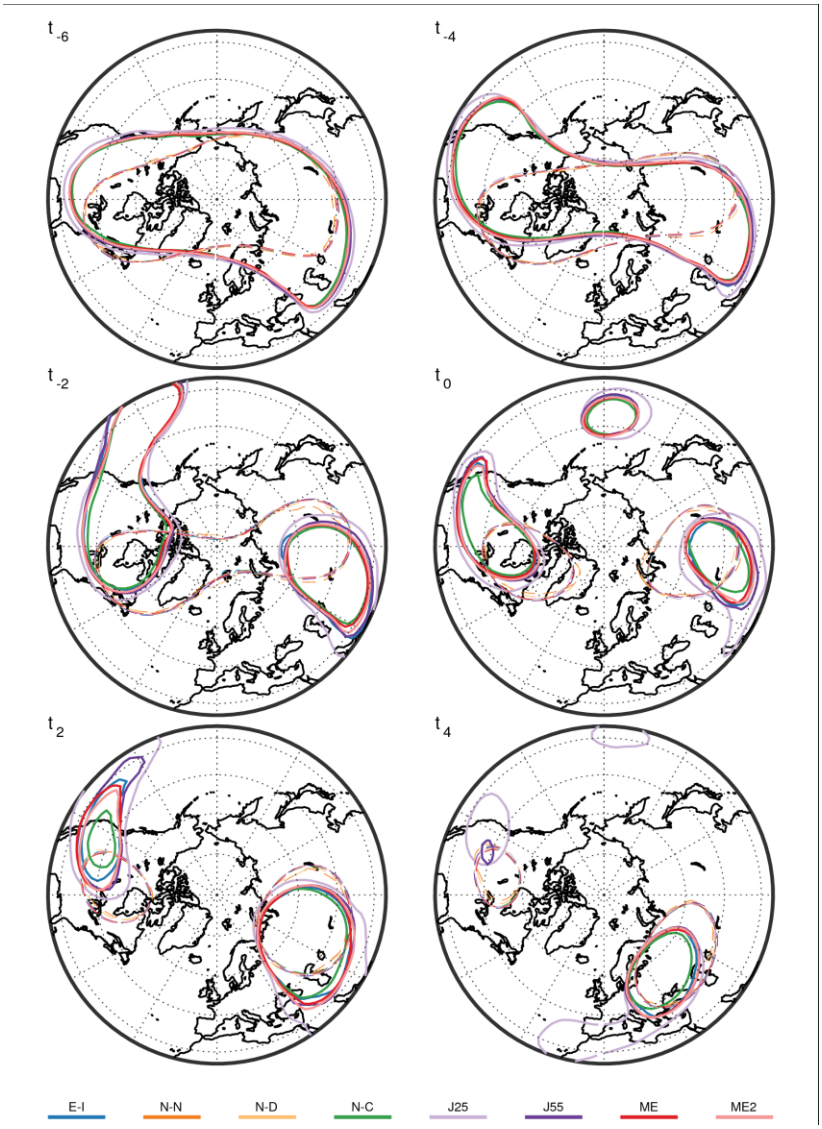


Figure 2: Vortex geometry at 10hPa (~~thin~~dashed) and 3 hPa (~~thick~~solid) during the 2009 SSW for different reanalysis datasets (colours). The 30km contour is illustrated at 10 hPa and the 38.5km contour is shown at 3 hPa. The central date (T_0) is set as January 24th 2009~~.~~

4 Evolution of zonal-mean flow and eddy fluxes

A composite analysis of the evolution of SSW events is first performed in Fig. 3 by showing the mean and standard deviation of various quantities averaged between 45 and 85°N. Only the LRE members are considered for this analysis. The evolution of zonal-mean wind highlights the rapid deceleration of the stratospheric polar vortex characterizing SSW events. At the same time, a warming of the stratosphere is observed. The agreement between reanalyses is generally good for both wind and temperature in the lower atmosphere. However, reanalyses show a large spread in $\bar{u}$ (~1 m~~/~~s~~-1~~) in the upper stratosphere, especially when winds are at their weakest. The spread in $\bar{T}$ also increases towards the upper stratosphere and tends to be larger (~5K) after the occurrence of SSW events. The zonal-mean meridional circulation ($\bar{v}$) becomes increasingly southward in the ~~high~~ upper stratosphere before the events, peaking a few days before the reversal of zonal-mean zonal wind. The Coriolis torque resulting from this circulation explains the deceleration of the vortex in the Eulerian framework (Matsuno, 1971). This is in contrast to the transformed Eulerian mean (TEM) framework where the residual circulation is poleward and downward during SSW events ~~(Song and Chun, 2016)~~(Limpasuvan et al., 2004). Whereas the agreement is good among reanalyses in the troposphere, larger spread is again observed in the upper stratosphere, coinciding with the minimum of $\bar{v}$.

The remainder of Fig. 3 illustrates the evolution of zonal-mean eddy heat ($\overline{v'T'}$) and momentum ($\overline{u'v'}$) fluxes, which are indicative of Rossby wave propagation. Here, both heat fluxes and momentum fluxes are clearly enhanced in the upper stratosphere before the central date, which corresponds to upward and equatorward wave propagation. Again, there is good agreement between reanalyses in the troposphere and lower stratosphere, whereas it is severely degraded in the upper stratosphere. The largest spread occurs coincidentally with the peaks in eddy fluxes and thus with the period when the most intense Rossby wave propagation occurs.

Composites of SSW events are not only subject to uncertainties related to the choice of dataset as shown in Fig. 3, but also to uncertainties related to the large diversity of events included in composites. Figure 4 first shows the standard deviation among SSW events for the same quantities shown in Fig. 3. Uncertainties related to the composite methodology are increasing in the upper stratosphere for all quantities~~.~~, and they also increase before the onset of SSW events (lag 0), which is the period when these terms have large magnitudes (compare with left column of Fig. 3). Zonal wind shows the largest uncertainty around lag 0 with a minimum at lag 0. This minimum results from the fact that SSW events were defined as a reversal of zonal-mean zonal wind at lag 0, which by construction forces all events to have similar zonal-mean zonal winds at 10 hPa.

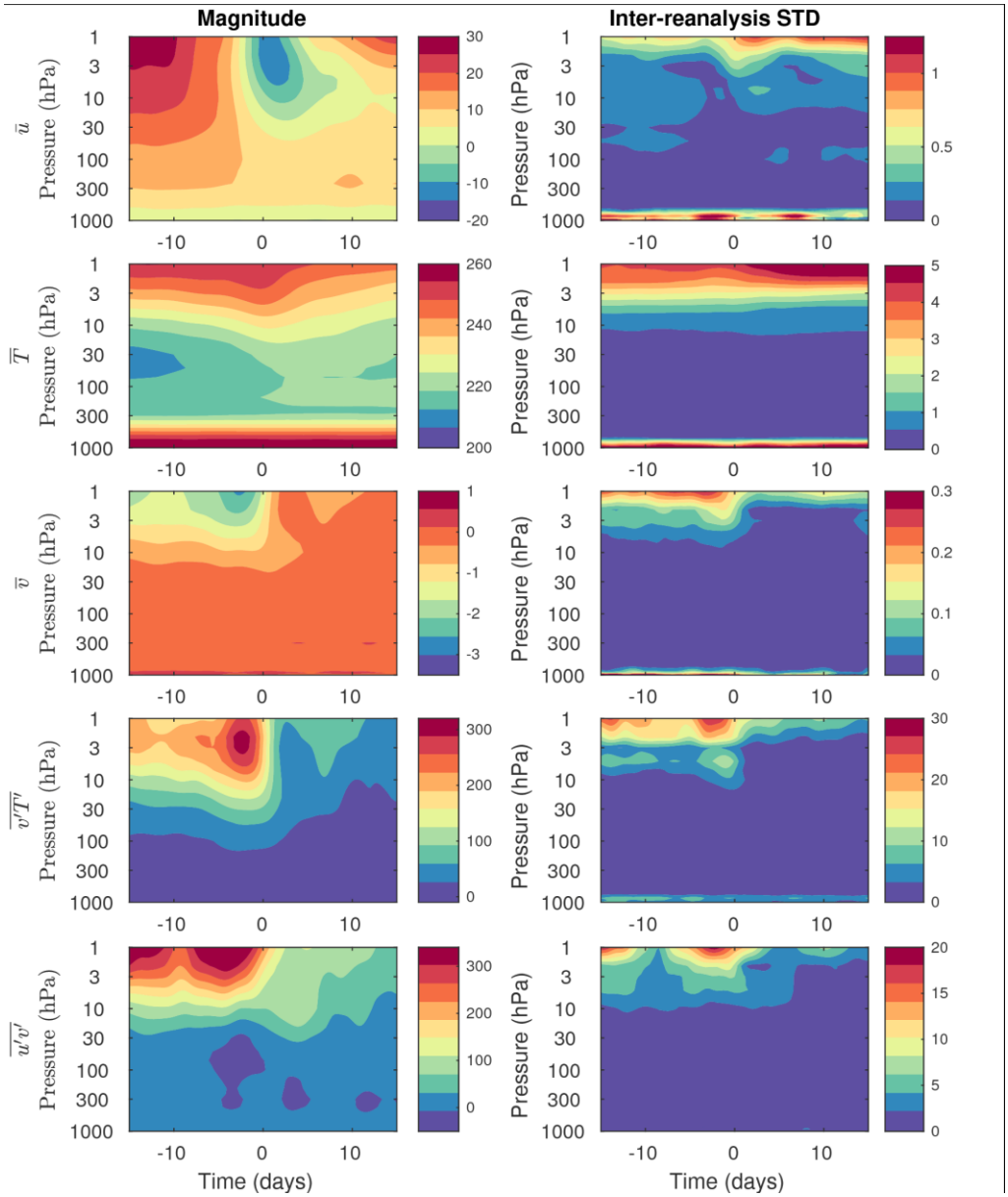


Figure 3: Evolution of zonal-mean variables and eddy fluxes during SSW events (rows) in function of pressure and time. All quantities are averaged from 45°N to 85°N. The LRE mean is shown to the left and the LRE standard deviation is shown to the right. Zonal wind ($\overline{u}$) and meridional wind ($\overline{v}$) have units of m/s^{-1}. Temperature ($\overline{T}$) has unit of K. Heat flux ($\overline{v'T'}$) has unit of mK/s. Momentum flux ($\overline{u'v'}$) has units of m^2/s^{-2}.

The ratio between the standard deviation among SSW events (Fig. 4 left column) and the standard deviation among datasets (Fig. 3 left column) is then shown in the right column of Fig. 4. This ratio is typically small in the upper troposphere and lower stratosphere, indicating that uncertainty is dominated by the large diversity of SSW events. Except zonal-mean zonal wind, most quantities show enhanced ratios in the upper stratosphere with values of about 0.15. Near the surface, many quantities also show large ratios. This suggests that the uncertainties related to the inter-reanalysis spread have ~~more importance to~~a larger impact on ~~our interpretation~~our understanding of the evolution of SSW events at the lower boundary and in the upper stratosphere.

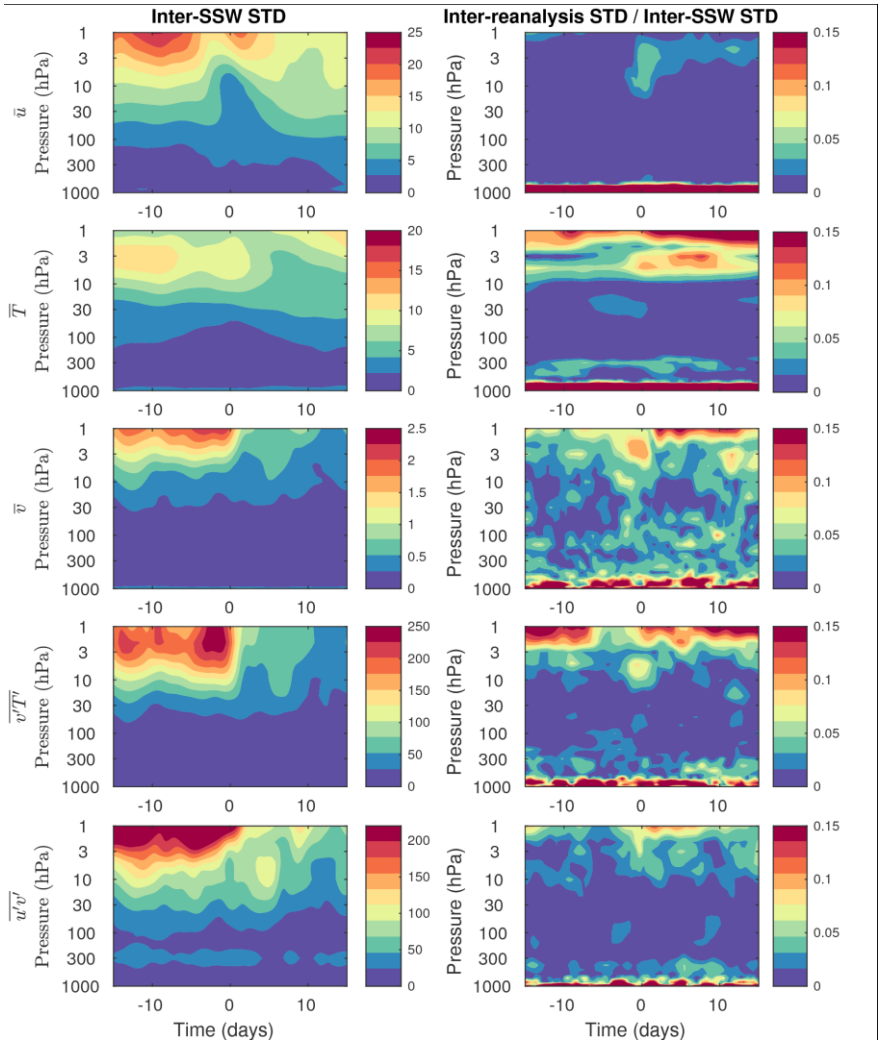


Figure 4: Similar to Fig. 3 except showing (left) the standard deviation among SSW events and (right) the standard deviation among reanalysis datasets divided by the standard deviation among SSW events.

The complete budget of zonal-mean momentum is then investigated in Fig. 5. The evolution of each term of Eq. (1.1) is illustrated at two representative levels in the stratosphere: 3 and 10 hPa. The

corresponding spread among reanalyses is shown using a logarithmic scale to allow the comparison of uncertainties over a wide range. Most SSW events are characterized by an intense deceleration of the zonal-mean zonal wind for a few days until the central date and immediately followed by a subsequent acceleration. Note the latter acceleration is weaker than the maximum deceleration during the onset, not reverting the zonal wind back to its original strength. The deceleration is most intense at 3 hPa but is also well observed in the mid-stratosphere at 10 hPa. Upper-tropospheric zonal-wind tendencies (not shown) are very weak in comparison to the stratosphere, and therefore almost undetectable if plotted on the same scale. The inter-reanalysis STD increases towards the central date where it peaks at 0.3 $ms^{-1}day^{-1}$ ~~m/s/day~~ at 3 hPa. The spread at 10 hPa is typically smaller and peaks around the onset date.

Moving on to the forcing terms of the momentum equation, the Coriolis torque ($f\bar{v}$), a term included in the QG scaling of the momentum equation, shows the largest forcing of all. This term is responsible for a large deceleration of the stratospheric vortex in the upper stratosphere. It peaks at about 2 to 3 days before the onset date and is markedly more intense in the upper stratosphere. Of all the resolved forcing terms, it also shows the largest spread among reanalyses, peaking slightly above 1 $ms^{-1}day^{-1}$ ~~m/s/day~~ at 1 to 2 days before the onset date. The uncertainty then decreases at the same time as the forcing itself becomes weaker in the upper stratosphere. In comparison to the upper stratosphere, the inter-reanalysis STD in the mid-stratosphere is smaller and more constant over time.

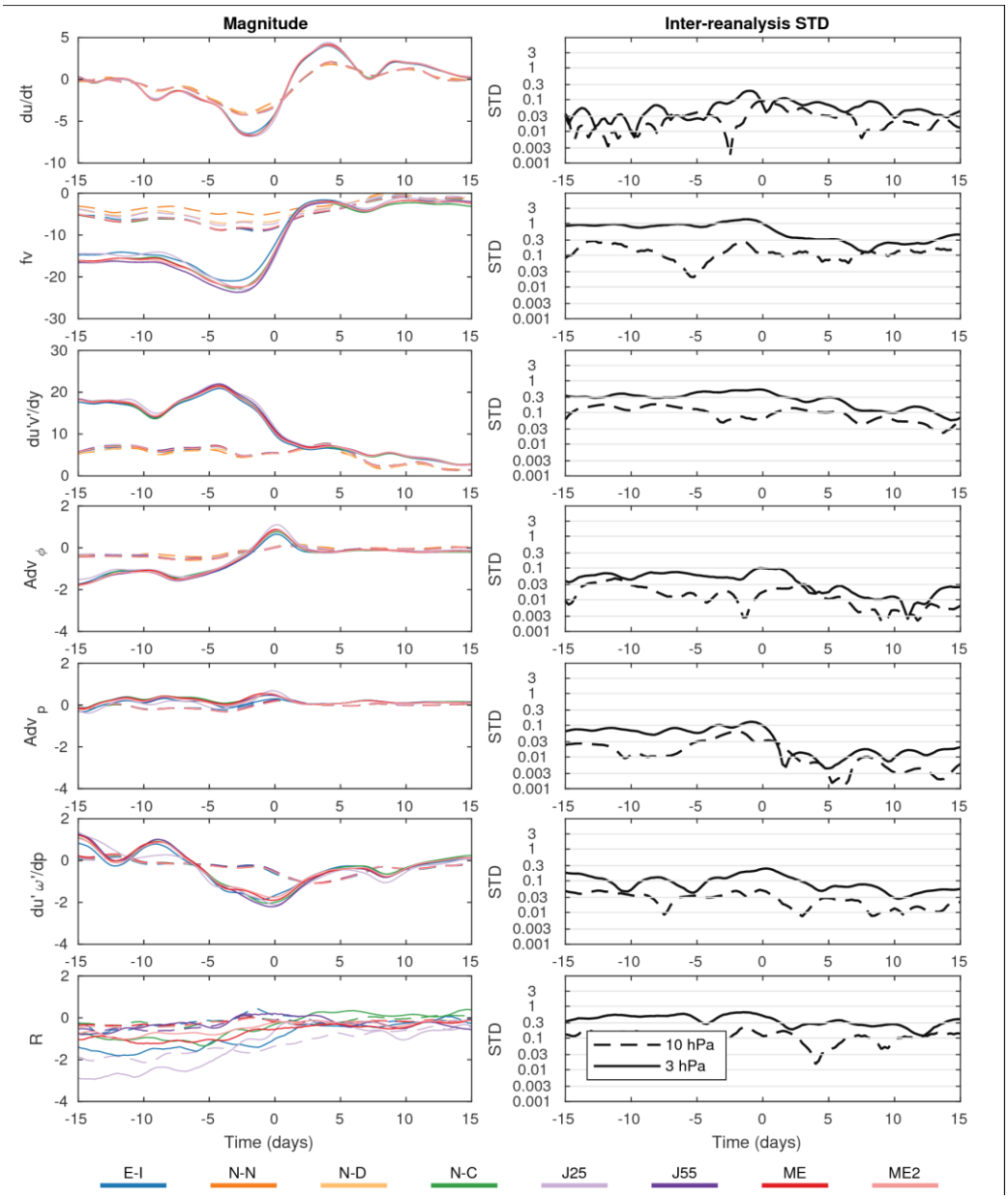


Figure 5: (left) Evolution of forcing terms of the zonal-mean momentum equation at 10 hPa (dashed lines) and 3 hPa (solid lines) in the course of SSW events. All variables are averaged from 45ºN to 85ºN. Note that the range of y-axis in each panel is different. (right) The inter-reanalysis spread (standard deviation) of the corresponding terms are shown for the LRE members. The standard deviation is shown on a logarithmic scale: the spacing between tick marks represents a decrease or increase of the standard deviation by a factor of about 3. All quantities are expressed in $ms^{-1}day^{-1}$ ~~m/s/day~~.

The momentum flux convergence, another term included in the QG scaling of the momentum equation, also shows large forcing in the upper stratosphere. It is largely opposed, but not completely, to the Coriolis torque. Similar to the Coriolis torque, inter-reanalysis STD peaks several days before the onset date (between 0.3 and 1 $ms^{-1}day^{-1}$~~m/s/day~~) and is reduced afterwards, especially in the upper stratosphere.

Terms that are left out of the QG form of the momentum equation (Fig. 5, rows 4 to 6) provide ~~much~~ smaller forcing for zonal wind tendency during SSW events in comparison to the convergence of horizontal momentum fluxes and the Coriolis torque. Their differences from one reanalysis to the other are also generally ~~much~~ smaller than those of QG terms. These terms are therefore not a large source of uncertainty in the momentum budget. It is nonetheless worth noting that the convergence of vertical fluxes of momentum is not negligible near the onset date of the event. Forcing magnitudes can reach up to about -2 $ms^{-1}day^{-1}$~~m/s/day~~. Its inter-reanalysis STD can also be relatively large (up to 0.3 $ms^{-1}day^{-1}$~~m/s/day~~) in the upper stratosphere, but is still small in comparison to the other dominant terms of the momentum equation.

The residual, which quantifies the consistency of the momentum diagnostics, is typically negative before the onset. This likely reflects the exclusion of gravity wave drag from the momentum budget (Martineau et al., 2016). Interestingly, its magnitude decreases in the upper stratosphere after SSW events (especially clear in JRA-25). This could be explained by the relatively quiet period following SSW events, when incoming fluxes of planetary-scale waves and gravity waves are supressed (Hitchcock and Shepherd, 2013). Similarly, the residual in the upper stratosphere exhibits a larger spread among reanalyses before the onset date in comparison to after.

The vertical dependence of the forcing terms and their inter-data spread are further investigated using vertical profiles averaged over five days before the onset date (Fig. 6). This period encompasses the period of large inter-reanalysis STD seen in the Coriolis force and momentum flux convergence (Fig. 5). All terms of the momentum equation, wind tendency, and forcing terms show increasingly large magnitudes in the upper stratosphere. Similarly, the inter-data STD is typically small in the troposphere but increases sharply in the stratosphere. Consistent with Fig. 5, terms with the largest inter-reanalysis STD include the Coriolis torque and the momentum flux convergence. Both show a noticeably improved agreement in the stratosphere by considering only the latest reanalysis ensemble (LRE) members instead of all reanalyses. We note that the inter-reanalysis STD becomes more similar between LRE and all datasets above 10 hPa where NCEP-NCAR and NCEP-DOE are left out in the all reanalysis ensemble due to the unavailability of data. This suggests that a substantial fraction of the inter-data STD below 10 hPa is attributable to these two reanalyses.

Compared to QG terms, non-QG terms are smaller in magnitude and agree better. Interestingly, the vertical convergence of momentum fluxes presents a sharp dipole in the vertical near the tropopause. This term, which involves a vertical derivative, may not be adequately resolved when computed with a

coarse vertical resolution. In fact, Martineau et al. (2016) have shown that using more vertical levels reduces the residual in the upper troposphere and lower stratosphere (see their Fig. A1b). In Fig. 6, the residual, R, becomes increasingly negative in the upper stratosphere. Again, this is largely a consequence of the exclusion of parameterized gravity wave drag from the momentum budget (Martineau et al., 2016).

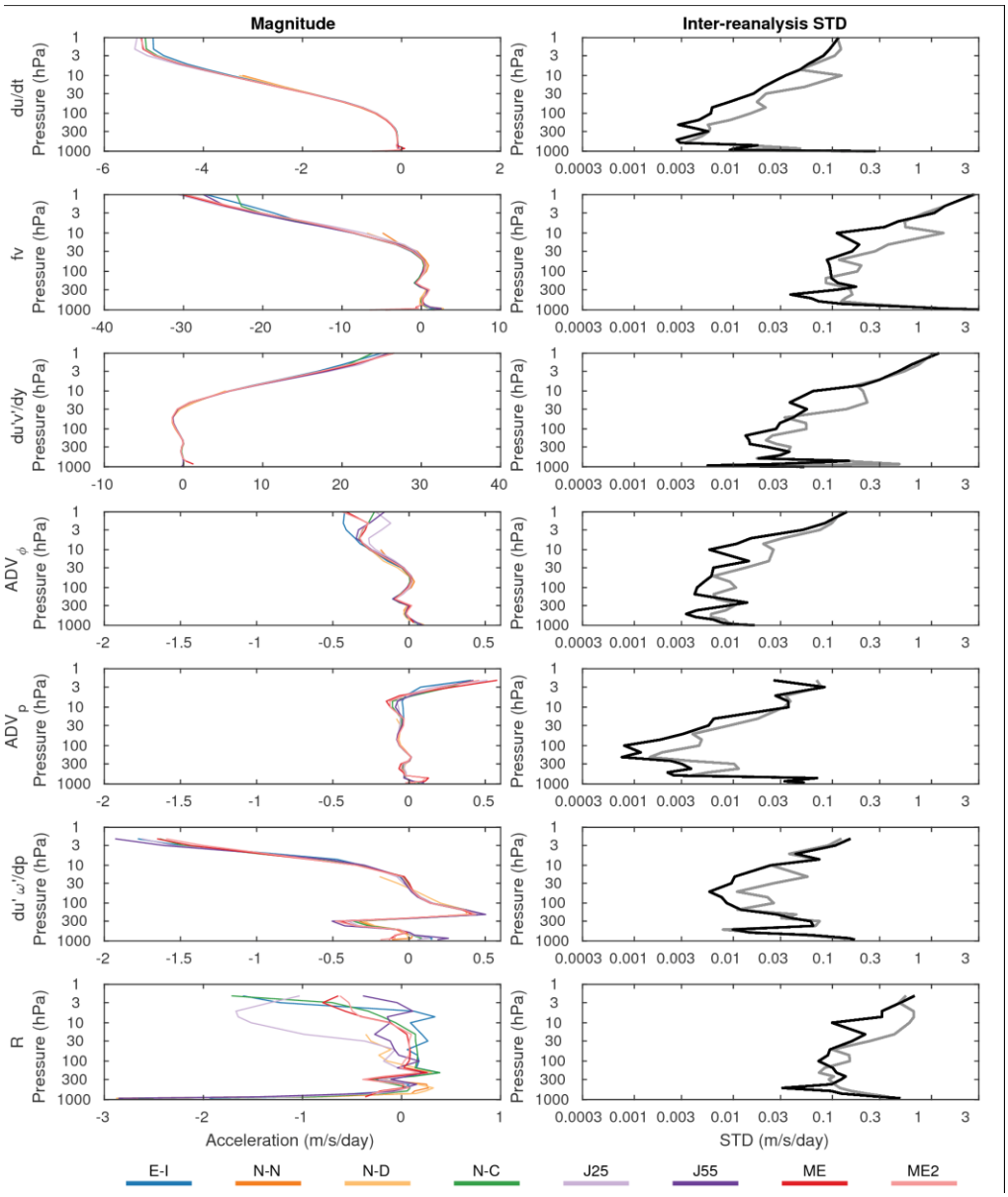


Figure 6: Vertical profiles of each term in the momentum equation averaged from lags -5 to 0 days during SSW events. All variables are averaged between 45°N and 85°N. Individual reanalyses are shown to the left and the inter-reanalysis standard deviation is shown to the right on a logarithmic scale. The latter is shown for all reanalyses (grey) and the LRE members (black). All quantities are expressed in units of ms^{-1}day^{-1} ~~m/s/day~~.

Each term of the momentum equation and the spread between reanalysis datasets are further evaluated as a function of latitude and pressure in Fig. 7 for the same period shown in Fig. 6 (day -5 to

0). The zonal wind deceleration during this period is characterized by a strong deceleration maximized at about 70°N. A weak acceleration is also present around 30°N in the upper stratosphere, resulting primarily from the Coriolis torque (Martineau and Son, 2015). Zonal wind tendency is fainter in the troposphere. As mentioned earlier, the agreement between datasets is much better in the lower atmosphere than the middle atmosphere. Large discrepancies are limited to the upper stratosphere and peak where the deceleration of the vortex is strongest.

The QG terms, as expected, are strongly opposed to each other in both the troposphere and the stratosphere. The Coriolis torque is responsible for most of the deceleration of the stratospheric vortex. As mentioned above, it is also responsible for zonal wind acceleration in the midlatitudes and subtropics. The efficiency of this forcing depends on the extent to which it is opposed by the convergence of fluxes of momentum. While they are strongly opposed in the upper troposphere, they are not perfectly balanced in the stratosphere, which results in the observed zonal wind tendencies. While the deceleration of the stratospheric vortex and its forcing is observed poleward of 50°N and thus well captured by averaging poleward of 45°N, QG forcing terms show a strong tripole in the troposphere, associated with the Hadley, Ferrel and polar cells, which results in some unavoidable cancellation of the forcing when averaging from 45 to 85°N, explaining why we do not observe large QG terms in the troposphere in Fig 6. The spread in the QG terms is maximized in the high-latitude stratosphere and a substantial reduction of the spread between datasets is evident in the lower stratosphere.

As expected, non-QG terms show ~~much~~ smaller forcing and inter-data spread in the high latitudes. Interestingly, advection terms are maximized in the subtropical upper stratosphere and tropopause. Among the non-QG terms, only the vertical convergence of momentum fluxes shows substantial forcing for deceleration in the upper stratosphere in the mid-latitudes. It also displays large and sharp forcing at the tropopause, a feature observed clearly in averages over the high-latitudes (Fig. 6). The resulting residual is typically negative and is also maximized in the high latitudes. On the other hand, it is larger in the jet's vicinity in the troposphere. The largest inter-data STD of the residual is in the polar region in the upper stratosphere.

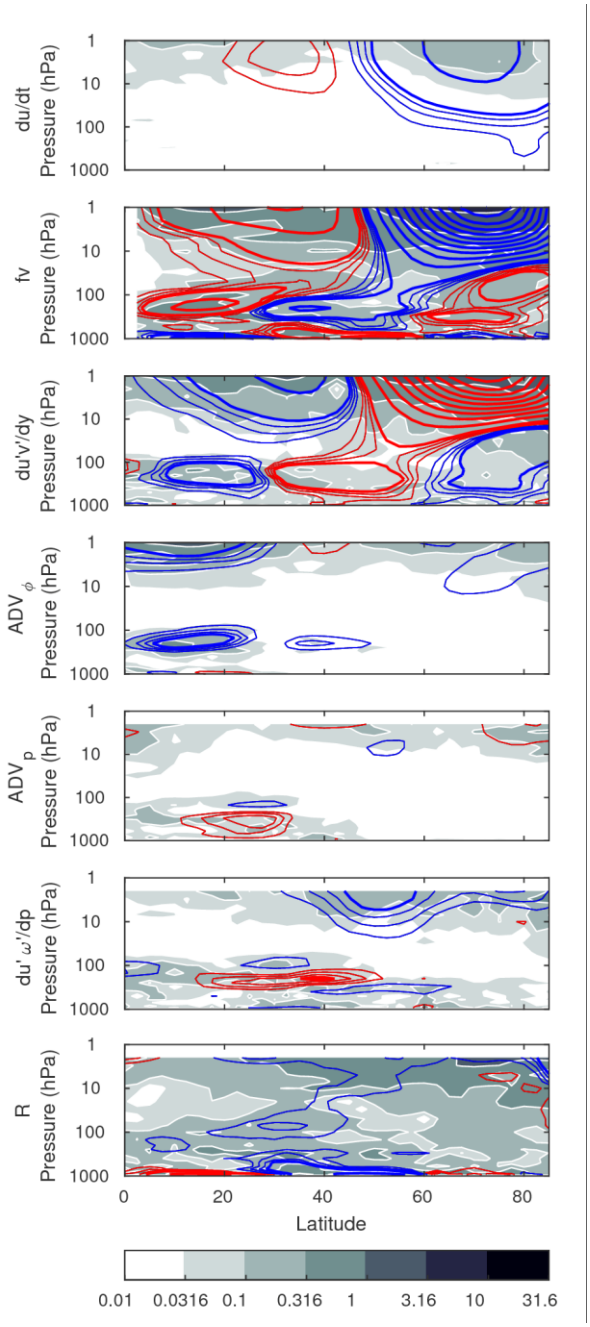


Figure 7: Pressure-latitude cross-section of inter-reanalysis mean terms of the momentum equation averaged from lags -5 to 0 days during SSW events. Composite mean is shown with red or blue for positive or negative values, with a contour interval of 0.25 for values ranging from 0.5 to 1.5 with thin contours and intervals of 2.5 ranging from 2.5 and larger with thick contours. The inter-reanalysis standard deviation is shaded per the colour bar. All quantities are shown for the LRE members and are expressed in units of $ms^{-1}day^{-1}$ ~~m/s/day~~.

The inter-dataset spread is further highlighted for QG terms in Figs. 8 and 9. The two figures show the difference between each reanalysis and the mean of the LRE members. The large-scale structure of the Coriolis torque (Fig. 8) is generally similar among reanalyses but notable differences are observed in the upper stratosphere. Particularly, both ERA-Interim and NCEP-CFSR show weaker forcing for deceleration (positive bias) in the high-latitude upper stratosphere in comparison to others. Looking at the same

cross-section for the momentum flux convergence (Fig. 9) it is noticeable that there is generally a better agreement with the mean of the LRE members in the lower stratosphere and troposphere. However, differences are still large in the upper stratosphere. Interestingly, the spatial distribution of biases in the Coriolis torque and the momentum flux convergence are somewhat opposed in the upper stratosphere in some reanalyses which may be indicative of a compensation between biases of the two leading terms of the momentum equation.

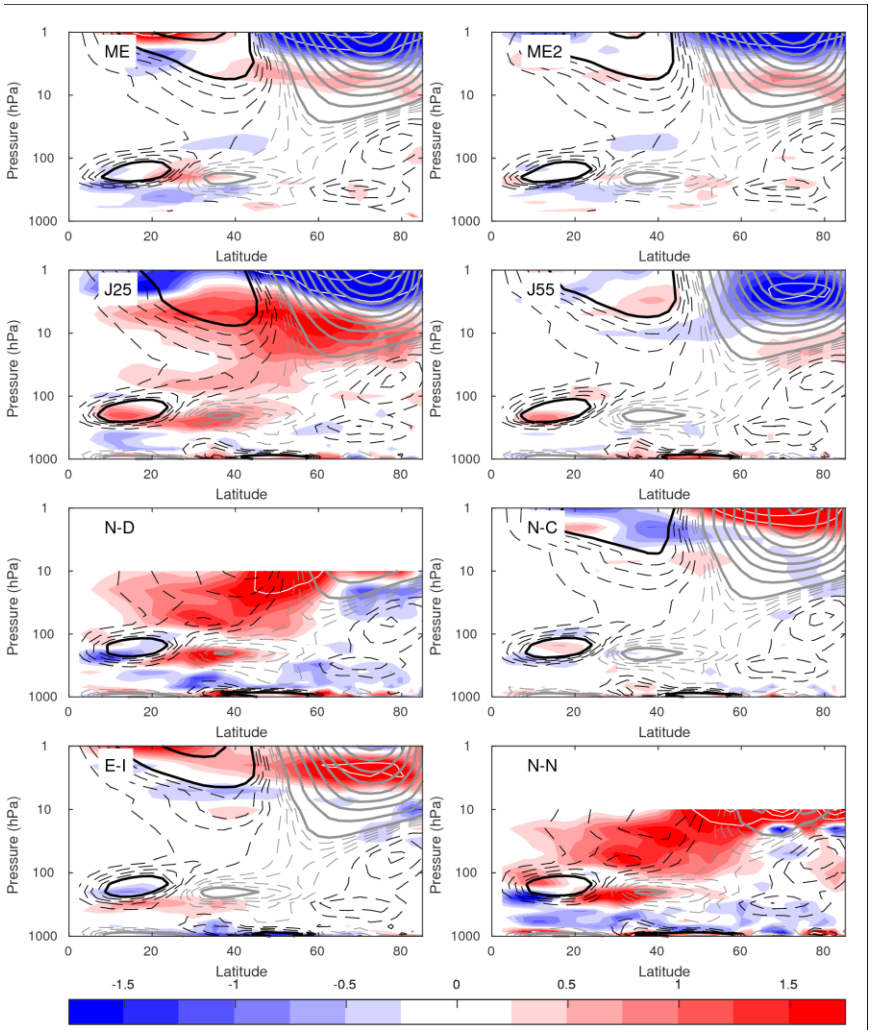


Figure 8: Latitude/pressure cross-section of the Coriolis torque averaged from lag -5 to 0 days during SSW events for different reanalysis datasets. The positive and negative values are contoured in black and grey, respectively. ~~Thin~~ Dashed contours range from -4 to 4 $ms^{-1}day^{-1}$ ~~m/s/day~~ in steps of 1 $ms^{-1}day^{-1}$ ~~m/s/day~~ and ~~thick~~ solid contours are used from ~~for~~ 5 $ms^{-1}day^{-1}$ ~~m/s/day~~ and larger in steps of 5 $ms^{-1}day^{-1}$~~m/s/day~~. Biases with respect to the mean of LRE members are illustrated with red and blue shading for positive and negative values, respectively. The colour interval is 0.25 $ms^{-1}day^{-1}$ ~~m/s/day~~ from -1.75 to 1.75. Values larger (smaller) than 2 (-2) are contoured every 2 $ms^{-1}day^{-1}$ ~~m/s/day~~ in white.

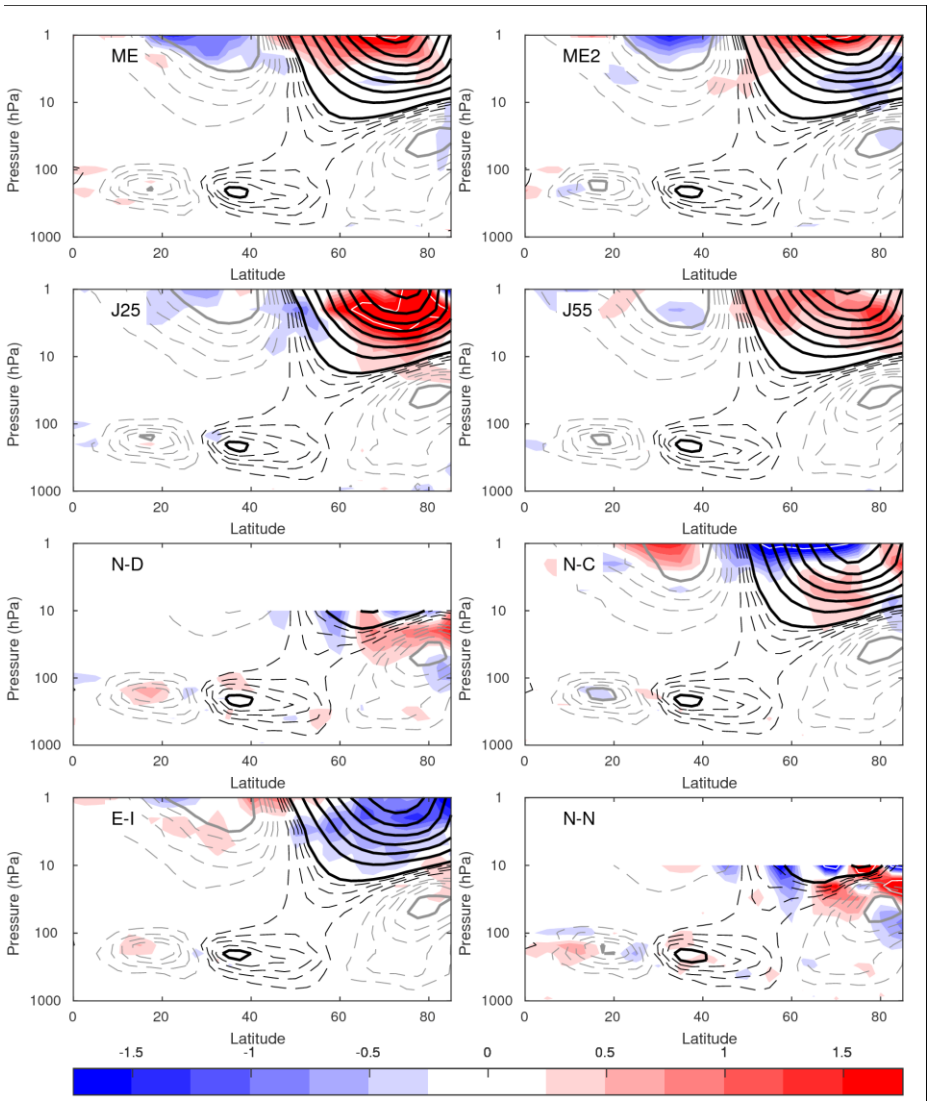


Figure 9: Same as Fig. 8 but for momentum flux convergence.

In order to identify possible sources of inconsistency in the momentum equation, the relationship between the residual and the dominant forcing terms is explored in Fig. 10 at three representative levels (3, 10 and 300 hPa). This analysis relies on the assumption that the inter-reanalysis variability of R is due to differences in the forcing terms and not to differences of zonal-wind tendency. This assumption is approximately valid since discrepancies in zonal wind tendencies are generally much smaller than discrepancies in the dominant terms of the momentum equation (see Figs. 5, 6 and 7). Similar to Martineau et al. (2016), a large fraction of the variability in the residual among reanalyses in the mid-stratosphere (10 hPa) can be attributed to the Coriolis torque (r=-0.99). Some relationship, although less significant, is also seen between the residual and the momentum flux convergence at 10 hPa (r=-0.80). Added together, the QG terms can explain all of the variability of the residual (r=-1). While this high correlation owes partly to the fact that JRA-25 is an outlier, NCEP-NCAR and NCEP-DOE also hint to a strong relationship between QG terms and the residual by inspecting the QG residual (R_{QG}), which is

computed by excluding non-QG terms from Eq. (1.1). The dominant role of the Coriolis torque is not as evident in the upper stratosphere (3 hPa) and the upper troposphere (300 hPa) but still plays an important role (r=-0.57 and r=-0.85, respectively).

The scatterplots highlight a convergence of newer reanalyses datasets in some circumstances. This is mostly apparent at 10 hPa where ERA-Interim, NCEP-CFSR, JRA-55, MERRA and MERRA2 are strongly clustered together. At 300 hPa, however, MERRA and MERRA2 tend to be apart from other reanalyses and at 3 hPa, JRA-55 and ERA-Interim tend to differ from the others. One should therefore not assume that there is always convergence when considering the latest datasets.

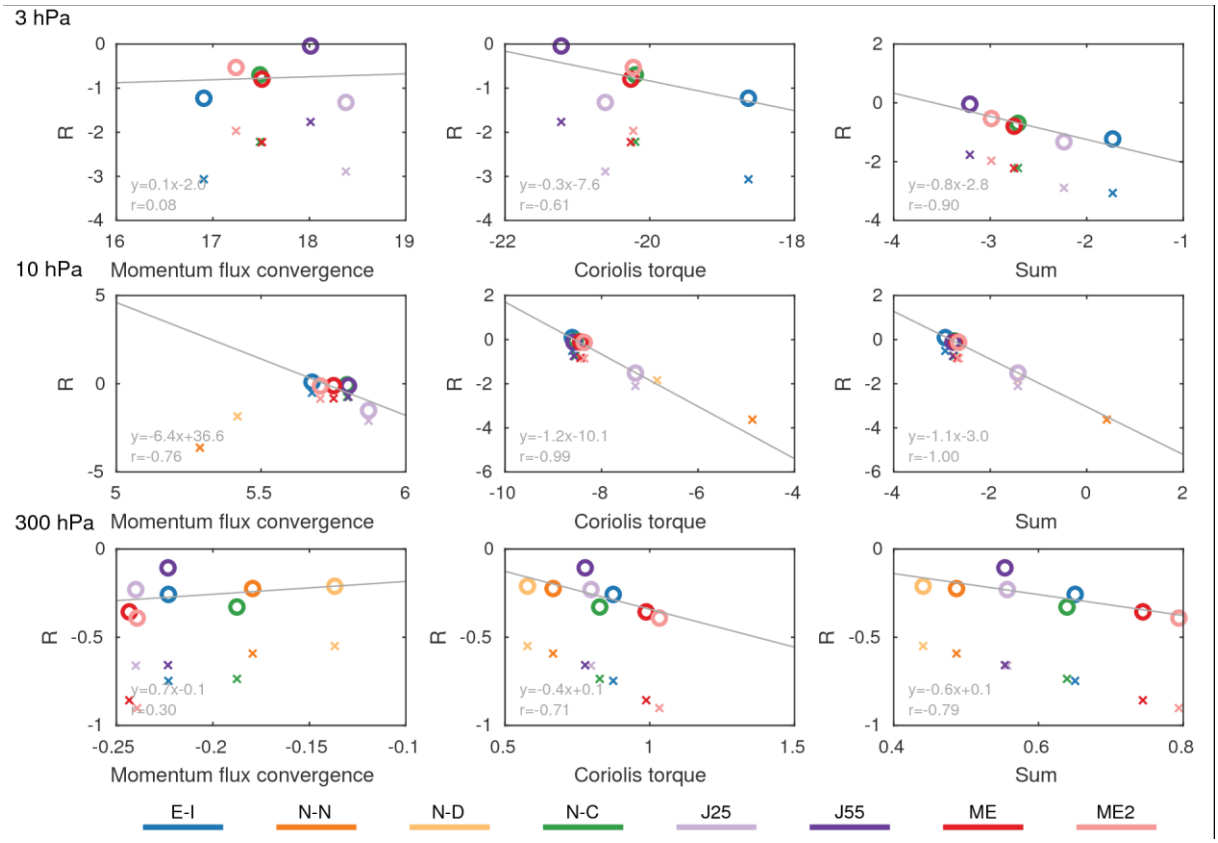


Figure 10: Scatterplot comparing the residual (R) with respect to (left) momentum flux convergence, (centre) Coriolis torque, and (right) their sum. While R is shown with circles, R_{QG} is shown using crosses. All variables are averaged between 45°N and 85°N and from lag -5 to 0 during SSW events. The (top) 3 hPa, (middle) 10 hPa, and (bottom) 300 hPa levels are shown. All variables are expressed in units of $ms^{-1}day^{-1}$ ~~m/s/day~~. Linear regression of R with respect to each forcing term is displayed in each panel.

To further explore a cancellation between the dominant terms of the momentum equation, the linear relationship between the Coriolis torque and the momentum flux convergence among reanalyses is also evaluated at the same three representative pressure levels (300, 10 and 3 hPa, not shown). While they are weakly opposed at 10 hPa and 300 hPa, there is a clear and strong anti-correlation between the two forcing terms at 3 hPa (r=-0.82). Reanalyses that show an enhanced deceleration by the Coriolis torque typically exhibit an enhanced forcing for acceleration by momentum flux divergence. Although the compensation is not perfect (slope of -1.3 when regressing Coriolis torque on momentum flux convergence), this compensation helps nonetheless to reduce the residual of the momentum equation, but does so at the expense of the accuracy of the forcing terms of the momentum equation.

To further identify causes of inter-reanalysis discrepancies, we now proceed to compare SSW events with high agreement between reanalyses (HASSWs) and those with low agreement (LASSWs). The classification of these events was described in Section 2.3. Figure 11 shows several momentum budget terms at 10 hPa during both event types. It is found that LASSWs are markedly more intense than HASSWs as seen by the peak deceleration of zonal wind that ~~is twice as strong~~stronger around the onset of the SSWs. The forcing terms are also markedly larger during LASSWs. Despite the larger forcing terms, a strong cancellation is observed between the Coriolis torque and momentum flux convergence. On the other hand, HASSWs show only steady and moderate forcing~~.~~. To investigate the role played by wave drag in both types of event, we illustrate EP flux convergence and the contribution of zonal wavenumber-1 planetary waves (see Appendix B of Martineau et al., 2016 for a formulation of EP flux and TEM momentum equation in pressure coordinates). We find that LASSWs present substantially larger convergence of wave activity fluxes in the stratosphere in comparison to HASSWs. Most of the difference in wave drag can be attributed to wavenumber-1 EP flux divergence.

Among the reanalyses, JRA-25 seems to stand out from the others in terms of EP flux divergence by wave-1 during LASSWs. Wave drag is especially affected by a positive bias at the early stage of these events. This difference could be due to a bias in stratospheric temperatures affecting the computation of heat fluxes and static stability, terms included in the computation of the vertical component of EP-flux. This bias is subsequently corrected in JRA-55 (Kobayashi et al., 2015), which could explain the better agreement between JRA-55 and other members of the LRE. Another visible outlier, ERA-Interim, underestimates wave drag during LASSW events. These discrepancies are observed in many events included in the composite (not shown).

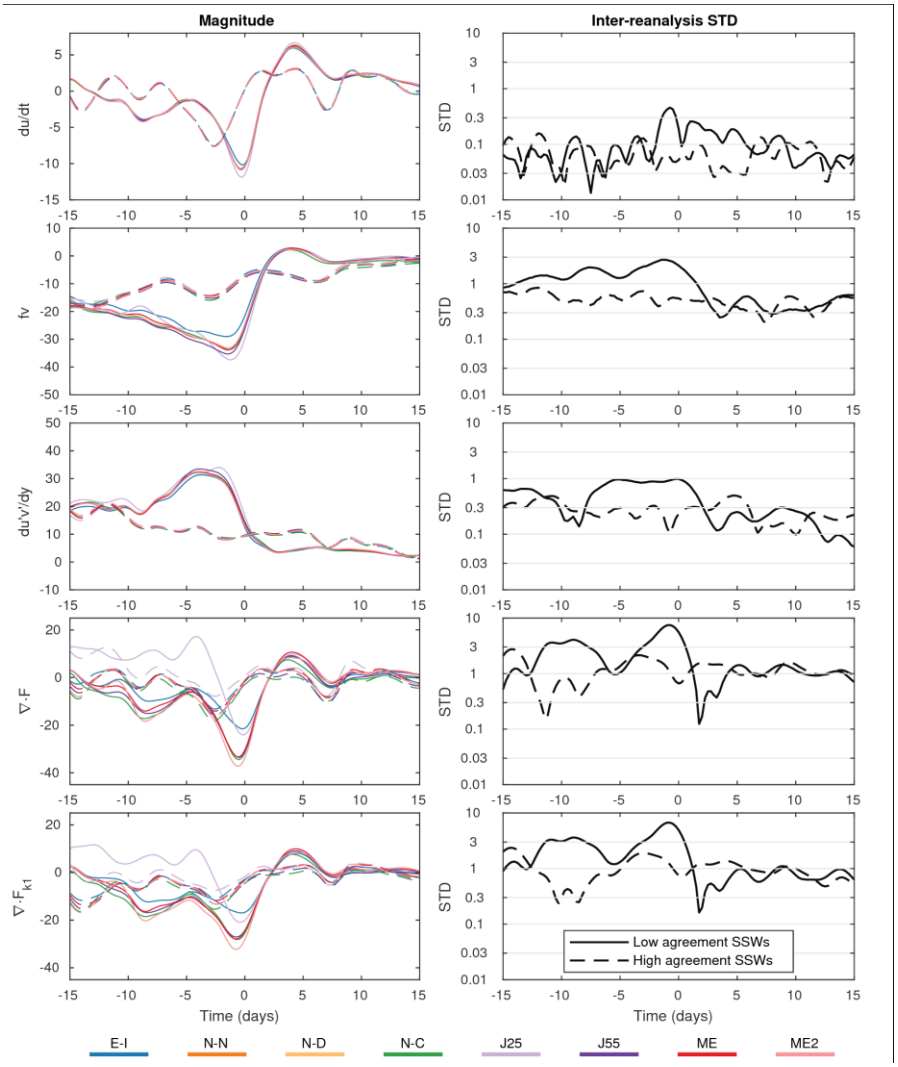


Figure 11: Similar to Fig. 5 but for comparing SSWs with small (HASSWs – dashed lines) and large uncertainties (LASSWs – solid lines). EP flux divergence is shown for all wave numbers (4th row) as well as wavenumber-1 (bottom row). All diagnostics are shown at 10 hPa and are in units of $ms^{-1}day^{-1}$ ~~**m/s/day**~~**.**

Next, the difference in the propagation of planetary-scale waves between HASSWs and LASSWs is illustrated in Fig. 12. LASSWs present markedly stronger forcing for deceleration (EP flux convergence) in the upper stratosphere in comparison to HASSW events and overall stronger inter-reanalysis uncertainty. Although LASSW events show both stronger upward and equatorward EP flux, most of the uncertainty is related to the vertical propagation alone. It is possible that these differences in wave drag among reanalyses be a source of large discrepancies in Coriolis torque through the generation of a compensating residual circulation ~~(Song and Chun, 2016)~~. However, as discussed earlier, the inter-reanalysis variability in momentum fluxes does not fully account for the variability in Coriolis torque (Fig. 11) in LASSWs. This suggests that other factors, such as gravity wave drag, data assimilation, and radiative transfer, could also play a role. Despite the larger discrepancies in stratospheric circulation

arising during LASSW events, there is no evidence of increased uncertainty of the tropospheric circulation in these events (not shown), indicating that either these differences have negligible impact on stratosphere-troposphere coupling or that any differences in downward coupling are prevented by data assimilation in the troposphere.

Since the previous analysis hinted that wave-1 fluxes of wave activity are an important source of uncertainty among reanalyses, we also compared displacement (SSWD) and split (SSWS) events to evaluate whether vortex geometry could be a source of uncertainty (not shown). The two types of SSWs show quite similar uncertainties among the various forcing terms. In fact, there is no clear separation between SSWD and SSWS; although SSWD events show more intense EP flux divergence by wavenumber-1, both types are forced by wavenumber-1 (Bancalá et al., 2012). Wavenumber-2 could simply show less uncertainty because it plays a lesser role in the composites shown. Among a total of seven LASSWs, four events are classified as SSWD events whereas three events are SSWS events and out of seven HASSWs four and two events are SSWDs and SSWSs, respectively. The remaining events are not clearly classified in either category. These statistics suffer however from the small sample sizes and there is no clear preference for split or displacement SSW events to be better represented in reanalysis datasets, as far as the inter-dataset agreement is concerned. On the other hand, when considering the dominant fluxes of wave activity producing SSW evens, we find that out of 7 LASSWs, four are W1-dominant and 1 is W2 dominant and out of 7 HASSWs, 1 is W1-dominant and 3 are W2-dominant. This seems to indicate that wavenumber-1 wave drag is responsible for larger uncertainties in reanalysis datasets but a detailed analysis reveals that inter-reanalysis spread is not markedly different between W1-dominant and W2-dominant events (not shown) suggesting that it is the intensity of wave drag rather than the longitudinal scale of wave activity that is linked to uncertainties among reanalyses.

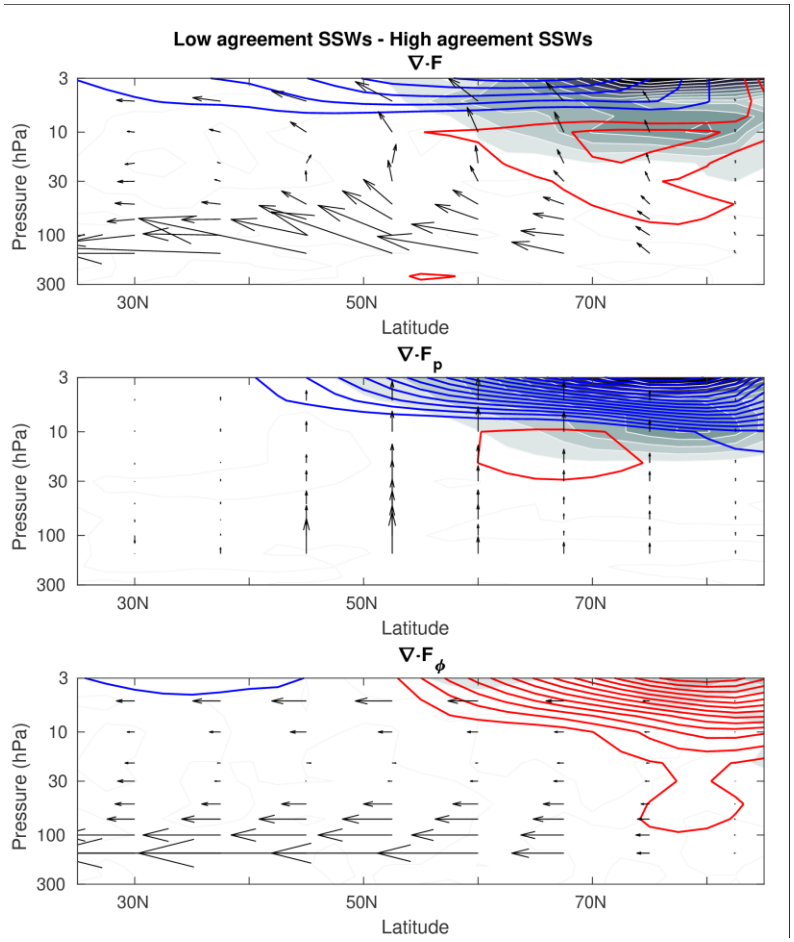


Figure 12: Difference in EP flux (arrows) and its divergence (contours) between HASSWs and LASSWs. All quantities are averaged over lags -7 to 2 days. EP flux convergence and divergence, respectively, are contoured in red and blue with an interval of 2.5 $ms^{-1}day^{-1}$~~m/s/day~~. Differences in inter-reanalysis spread are shown in grey with a shading interval of 1 $ms^{-1}day^{-1}$~~m/s/day~~. Only LRE members are considered in this figure.

5 Summary and conclusions

To assess uncertainties in the dynamical variability of the stratosphere and troposphere and their coupling in reanalysis datasets, a detailed comparison of zonal-mean momentum diagnostics is carried out for eight reanalysis datasets during sudden stratospheric warming events (SSWs). Emphasis is placed on the vertical and temporal dependence of the uncertainties during the events as well as on the factors that lead to the uncertainties.

From the troposphere to the mid-stratosphere, all quantities of the momentum equation are remarkably similar among the datasets. Although inter-data discrepancies increase substantially towards the upper stratosphere, zonal-mean zonal wind and temperature, often used to illustrate the vertical coupling during SSW events, agree quite well up to the mid-stratosphere. As such, the temporal-spatial evolution of composite SSW events is nearly identical in the different datasets (Martineau and Son, 2010; Palmeiro et al., 2015).

Non-negligible uncertainties are observed mainly in the upper stratosphere. They are particularly large during the most intense SSW events, indicating that uncertainties in the momentum tendency and the related eddy fluxes are related to the strength of the episodes of planetary-scale wave propagation from the troposphere to the stratosphere. No significant difference among reanalyses, however, is found when comparing splitting and displacement events.

Among all resolved forcing terms of the zonal-mean momentum equation, the Coriolis torque and the meridional convergence of momentum fluxes show the largest magnitudes and largest disagreement among reanalyses several days before the reversal of zonal wind in the stratosphere. Such uncertainties decrease dramatically after the zonal-mean winds change direction during SSW events. This could be explained by the fact that a dynamically quiet period generally follows SSW events (Hitchcock and Shepherd, 2013). Other forcing terms, i.e., non-QG terms, show smaller magnitude, and smaller inter-dataset uncertainty in comparison to the Coriolis torque and the convergence of momentum fluxes.

The large variability of forcing terms in the ~~high~~ upper stratosphere among reanalyses exceeds many times the uncertainties in zonal wind. Thus, some reanalysis datasets exhibit large residuals in the momentum equation. Interestingly, the residuals are large and vary substantially among datasets prior to the reversal of zonal wind. This is consistent with the enhanced residual observed in periods of vortex transience in the analysis of Martineau et al. (2016). A marked reduction of the magnitude of the residual after the reversal of the zonal-mean circulation again suggests a less dynamically active period where the residual may be decreased in part because of a reduction of gravity wave drag in the upper stratosphere (Hitchcock and Shepherd, 2013), which is not considered in our budget of momentum and thus included in the residual. Most of the residual in the stratosphere is correlated to uncertainties in the Coriolis torque which may indicate that the residual circulation responds in a way as to balance missing forcing, likely by unresolved wave drag, in the momentum equation. The same also holds in the troposphere. Unlike the mid-stratosphere and troposphere, however, the residual of the zonal-mean momentum equation in the upper stratosphere benefits from a cancellation between biases in the Coriolis torque and eddy momentum flux convergence. Although this phenomenon contributes to a seemingly improved momentum budget, it does not help to reduce the uncertainties in the dynamical evolution of the events. This uncertainty does not, however, overly alter our interpretation of the dynamics regulating SSW events and should therefore not be a concern when studying SSW events in the troposphere and stratosphere. This relationship also indicates that a fraction of the variability of the meridional circulation among reanalyses is driven by inter-dataset differences of eddy fluxes. Wave drag is, however, not the sole contributor to inter-reanalysis discrepancies. Biases in the mean state combined with data assimilation may lead to discrepancies in meridional circulation in the stratosphere (Kobayashi and Iwasaki, 2016; Uppala et al., 2005).

A reduction of the inter-reanalysis spread during SSW events is generally observed in newer reanalyses for most terms of the momentum equation, especially in the stratosphere. Inspection of individual

reanalyses however reveals that newer reanalyses are not always clustered together and that outliers vary depending on the pressure level and the variables.

Since this analysis compares the momentum diagnostics among reanalysis datasets, the uncertainties discussed here are resulting from both the evolution of the different atmospheric fields by the forecast process and their subsequent adjustment by the data assimilation process. Discrepancies among reanalyses may thus originate from differences in the models, observations being assimilated, and assimilation techniques. Sources of error also include processes that are not well captured or parameterized, the latter not being considered in the momentum budget of this study. As discussed in Martineau et al. (2016), some fields may be easier to constrain than others, and thus more representative of the true evolution of the atmosphere. For instance, the zonal-mean zonal wind may be better constrained by temperature observations by assumptions of balance, like thermal wind balance. On the other hand, ageostrophic flows such as the meridional circulation, may be harder to constrain by data assimilation. It is possible that this loosely-constrained circulation may act to oppose biases in other forcings. A better understanding of the distinct contributions of the modelling and data assimilations steps to the observed uncertainties would require studying the forecasts and analysis increments separately.

References

Abalos, M., Legras, B., Ploeger, F. and Randel, W. J.: Evaluating the advective Brewer-Dobson circulation in three reanalyses for the period 1979-2012, J. Geophys. Res. Atmos., 120(15), 7534–7554, doi:10.1002/2015JD023182, 2015.

Ayarzaguena, B., Langematz, U. and Serrano, E.: Tropospheric forcing of the stratosphere: A comparative study of the two different major stratospheric warmings in 2009 and 2010, J. Geophys. Res. Atmos., 116(18), 1–15, doi:10.1029/2010JD015023, 2011.

Baldwin, M. P.: Stratospheric Harbingers of Anomalous Weather Regimes, Science (80-.)., 294(5542), 581–584, doi:10.1126/science.1063315, 2001.

Bancalá, S., Krüger, K. and Giorgetta, M.: The preconditioning of major sudden stratospheric warmings, J. Geophys. Res. Atmos., 117(4), 1–12, doi:10.1029/2011JD016769, 2012.

Butler, A. H., Seidel, D. J., Hardiman, S. C., Butchart, N., Birner, T. and Match, A.: Defining Sudden Stratospheric Warmings, Bull. Am. Meteorol. Soc., 96(11), 1913–1928, doi:10.1175/BAMS-D-13-00173.1, 2015.

Charlton, A. J. and Polvani, L. M.: A New Look at Stratospheric Sudden Warmings. Part I: Climatology and Modeling Benchmarks, J. Clim., 20(3), 449–469, doi:10.1175/JCLI3996.1, 2007.

Compo, G. P., Whitaker, J. S., Sardeshmukh, P. D., Matsui, N., Allan, R. J., Yin, X., E. Gleason, J., 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., andT. F. Ross, Ø. N., Trigo, R. M., Wang, X. L., Woodruff, S. D. and Worleyu, S. J.: The Twentieth Century Reanalysis Project, Q. J. R. Meteorol. Soc., 137, 1–28, doi: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. and Bauer, P.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, Q. J. R. Meteorol. Soc., 137, 553–597, doi:10.1002/qj.828, 2011.

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(2), 1417–1452, doi:10.5194/acp-17-1417-2017, 2017.

Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., Randles, C., Darmenov, A., Bosilovich, M. G., Reichle, R., Wargan, K., Coy, L., Cullather, R., Draper, C., Akella, S., Buchard, V., Conaty, A., da Silva, A., 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., 2, JCLI-D-16-0758.1, doi:10.1175/JCLI-D-16-0758.1, 2017.

Harada, Y., Goto, A., Hasegawa, H., Fujikawa, N., Naoe, H. and Hirooka, T.: A Major Stratospheric Sudden Warming Event in January 2009, J. Atmos. Sci., 67, 2052–2069, doi:10.1175/2009JAS3320.1, 2010.

Hitchcock, P. and Shepherd, T. G.: Zonal-Mean Dynamics of Extended Recoveries from Stratospheric Sudden Warmings, J. Atmos. Sci., 70(2), 688–707, doi:10.1175/JAS-D-12-0111.1, 2013.

Iwasaki, T., Hamada, H. and Miyazaki, K.: Comparisons of Brewer-Dobson Circulations Diagnosed from Reanalyses, J. Meteorol. Soc. Japan, 87(6), 997–1006, doi:10.2151/jmsj.87.997, 2009.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K. C., Ropelewski, C., Wang, J., Leetmaa, A., Reynolds, R., Jenne, R. and Joseph, D.: The NCEP/NCAR 40-Year Reanalysis Project, Bull. Am. Meteorol. Soc., 77(3), 437–471, 1996.

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

Kidston, J., Scaife, A. A., Hardiman, S. C., Mitchell, D. M., Butchart, N., Baldwin, M. P. and Gray, L. J.: Stratospheric influence on tropospheric jet streams, storm tracks and surface weather, Nat. Geosci., 8(6), 433–440, doi:10.1038/ngeo2424, 2015.

Kobayashi, C. and Iwasaki, T.: Brewer-Dobson circulation diagnosed from JRA-55, J. Geophys. Res. Atmos., 121(4), 1493–1510, doi:10.1002/2015JD023476, 2016.

Kobayashi, S., Ota, Y., Harada, Y., Ebita, A., Moriya, M., Onoda, H., Onogi, K., Kamahori, H., Kobayashi, C., Endo, H., Miyaoka, K. and Takahashi, K.: The JRA-55 Reanalysis: General Specifications and Basic Characteristics, J. Meteorol. Soc. Japan. Ser. II, 93(1), 5–48, doi:10.2151/jmsj.2015-001, 2015.

Lehtonen, I. and Karpechko, A. Y.: Observed and modeled tropospheric cold anomalies associated with sudden stratospheric warmings, J. Geophys. Res. Atmos., 121(4), 1591–1610,

doi:10.1002/2015JD023860, 2016.

Limpasuvan, V., Thompson, D. W. J. and Hartmann, D. L.: The Life Cycle of the Northern Hemisphere Sudden Stratospheric Warmings, J. Clim., 17(13), 2584–2596, doi:10.1175/1520-0442(2004)017<2584:TLCOTN>2.0.CO;2, 2004.

Lu, H., Bracegirdle, T. J., Phillips, T. and Turner, J.: A Comparative Study of Wave Forcing Derived from the ERA-40 and ERA-Interim Reanalysis Datasets, J. Clim., 28, 2291–2311, 2015.

Manney, G. L., Schwartz, M. J., Kruger, K., Santee, M. L., Pawson, S., Lee, J. N., Daffer, W. H. and Fuller, R. A.: Aura Microwave Limb Sounder observations of dynamics and transport during the record-breaking 2009 Arctic stratospheric major warming, Geophys. Res. Lett., 36(L12815), doi:10.1029/2009GL038586, 2009.

Martineau, P. and Son, S.-W.: Quality of reanalysis data during stratospheric vortex weakening and intensification events, Geophys. Res. Lett., 37(L22801), doi:10.1029/2010GL045237, 2010.

Martineau, P. and Son, S.-W.: Planetary-scale wave activity as a source of varying tropospheric response to stratospheric sudden warming events: A case study, J. Geophys. Res. Atm., 118, doi:10.1002/jgrd.50871, 2013.

Martineau, P. and Son, S.-W.: Onset of Circulation Anomalies during Stratospheric Vortex Weakening Events: The Role of Planetary-Scale Waves, J. Clim., 28(18), 7347–7370, doi:10.1175/JCLI-D-14-00478.1, 2015.

Martineau, P., Son, S.-W. and Taguchi, M.: Dynamical Consistency of Reanalysis Datasets in the Extratropical Stratosphere, J. Clim., 29(8), 3057–3074, doi:10.1175/JCLI-D-15-0469.1, 2016.

Matsuno, T.: A Dynamical Model of the Stratospheric Sudden Warming, J. Atmos. Sci., 28, 1479–1494, 1971.

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(8), 2668–2682, doi:10.1175/JCLI-D-12-00030.1, 2013.

Monge-Sanz, B. M., Chipperfield, M. P., Dee, D. P., Simmons, a. J. and Uppala, S. M.: Improvements in the stratospheric transport achieved by a chemistry transport model with ECMWF (re)analyses: identifying effects and remaining challenges, Q. J. R. Meteorol. Soc., 139(672), 654–673, doi:10.1002/qj.1996, 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, 2007.

Palmeiro, F. M., Barriopedro, D., García-Herrera, R. and Calvo, N.: Comparing Sudden Stratospheric Warming Definitions in Reanalysis Data, J. Clim., 150709110342004, doi:10.1175/JCLI-D-15-0004.1, 2015.

Poli, P., Hersbach, H., Tan, D., Dee, D., Thépaut, J., Simmons, A., Peubey, C., Laloyaux, P., Komori, T., Berrisford, P. and Dragani, R.: The data assimilation system and initial performance evaluation of the

ECMWF pilot reanalysis of the 20th-century assimilating surface observations only (ERA-20C), ERA Rep. Ser., 59, 2013.

Polvani, L. M. and Waugh, D. W.: Upward Wave Activity Flux as a Precursor to Extreme Stratospheric Events and Subsequent Anomalous Surface Weather Regimes, J. Clim., 17(18), 3548–3554, doi:10.1175/1520-0442(2004)017<3548:UWAFAA>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(14), 3624–3648, 2011.

Saha, S., Moorthi, S., Pan, H.-L., Wu, X., Wang, J., Nadiga, S., Tripp, P., Kistler, R., Woollen, J., Behringer, D., Liu, H., Stokes, D., Grumbine, R., Gayno, G., Wang, J., Hou, Y.-T., Chuang, H.-Y., Juang, H.-M. H., Sela, J., Iredell, M., Treadon, R., Kleist, D., 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. Am. Meteorol. Soc., 91(8), 1015–1057, doi:10.1175/2010BAMS3001.1, 2010.

Saha, S., Moorthi, S., Wu, X., Wang, J., Nadiga, S., Tripp, P., Behringer, D., Hou, Y. T., Chuang, H. Y., Iredell, M., Ek, M., Meng, J., Yang, R., Mendez, M. P., Van Den Dool, H., Zhang, Q., Wang, W., Chen, M. and Becker, E.: The NCEP climate forecast system version 2, J. Clim., 27(6), 2185–2208, doi:10.1175/JCLI-D-12-00823.1, 2014.

Seviour, W. J. M., Mitchell, D. M. and Gray, L. J.: A practical method to identify displaced and split stratospheric polar vortex events, Geophys. Res. Lett., 40(2), 5268–5273, doi:10.1002/grl.50927, 2013.

Sigmond, M., Scinocca, J. F., Kharin, V. V. and Shepherd, T. G.: Enhanced seasonal forecast skill following stratospheric sudden warmings, Nat. Geosci., 6(1), 1–5, doi:10.1038/ngeo1698, 2013.

Smith, A. K. and Lyjak, L. V.: An Observational Estimate of Gravity Wave Drag From the Momentum Balance in the Middle Atmosphere, J. Geophys. Res., 90, 2233–2241, 1985.

Smith, K. L. and Kushner, P. J.: Linear interference and the initiation of extratropical stratosphere-troposphere interactions, J. Geophys. Res., 117(D13), D13107, doi:10.1029/2012JD017587, 2012.

Song, B. and Chun, H.: Residual Mean Circulation and Temperature Changes during the Evolution of Stratospheric Sudden Warming Revealed in MERRA, Atmos. Chem. Phys. Discuss., 17(November), 1–26, doi:10.5194/acp-2016-729, 2016.

Taguchi, M.: On the Asymmetry of Forecast Errors in the Northern Winter Stratosphere between Vortex Weakening and Strengthening Conditions, J. Meteorol. Soc. Japan. Ser. II, 93(4), 443–457, doi:10.2151/jmsj.2015-029, 2015.

Tripathi, O. P., Baldwin, M., Charlton-Perez, A., Charron, M., Eckermann, S. D., Gerber, E., Harrison, R. G., Jackson, D. R., Kim, B.-M., Kuroda, Y., Lang, A., Mahmood, S., Mizuta, R., Roff, G., Sigmond, M. and Son, S.-W.: The predictability of the extratropical stratosphere on monthly time-scales and its impact on the

skill of tropospheric forecasts, Q. J. R. Meteorol. Soc., 141(689), 987–1003, doi:10.1002/qj.2432, 2015.

Uppala, S. M., Kallberg, 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. M., Berg, L. Van De, Bidlot, J., Bormann, N., Caires, S., Chevallier, F., Dethof, A., Dragosavac, M., Fisher, M., Fuentes, M., Hagemann, S., Holm, E., Hoskins, B. J., Isaken, 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 re-analysis, Q. J. R. Meteorol. Soc., 131, 2961–3012, 2005.