Review of “Reanalysis intercomparison of potential vorticity and potential-vorticity-based diagnostics” by Millán et al.

Millán and colleagues study the differences of potential vorticity (PV) and PV-based diagnostics in four modern reanalyses (ERA-Interim, MERRA-2, JRA-55 and CFSR/CFSv2). The discussion centers around (i) the calculation of PV and the differences arising in this task in the various reanalyses, (ii) the impact of the data assimilation on PV in each reanalysis product, (iii) seasonal and annual mean variability of sPV between the various reanalyses, as well as of PV-based diagnostics such as (iv) equivalent latitude, (v) dynamic tropopause and (vi) polar vortex characterization. The major finding is that PV agrees well between the various data sets on the time scales studied in this work. The authors also highlight the situation where more caution is necessary when working with PV. Some differences between the various data sets arise in particular for (i) equivalent latitude calculations at low latitudes or high altitudes, (ii) the dynamic tropopause in regions of jetstreams and of strong topography, as well as (iii) during the formation and demise of the polar vortex.

This work is intended as part of the S-RIP special issue where it perfectly fits. Such a comparison of PV from different reanalysis data sets has not been completed yet, although PV from reanalysis is a widely used diagnostic to analyze transport and dynamics in the troposphere and stratosphere. The questions asked in the paper are clear. The analysis is very convincing; data and methods are well described. The figures are well structured and clear to understand. The conclusions are based on the analysis. It is easy to follow the thoughts of the authors and as a reader I have the feeling that the authors really know what they are talking about. In general I think this study will be of great value to users of reanalysis data and as such I would support publication of this study in ACP in the S-RIP special issue. I have some rather minor comments listed below, which the authors might consider for a revised version.

Comments:

- P3, L11: To my knowledge, ERA-I provides relative vorticity, see eg. the ERA-Interim data catalogue: https://apps.ecmwf.int/archive-catalogue/?stream=oper&levtype=ml&expver=1&month=jan&year=1979&type=an&class=ei

- Equation 1: maybe it is worth mentioning that this is the synoptic approximation of PV

- Equation 2: I think, since not everybody might be familiar with sPV, some readers would benefit from a comparison of sPV and PV, maybe shown here for one reanalysis but for different averaging times, e.g., a snapshot, monthly, or yearly mean.

- P3, l26: Could you mention the potential temperature range here.

- P4, l34: Could you say something about the cause of the low bias of sPV? Is this an effect of vertical grid spacing or model physics (e.g. GW drag)?

- Sec. 4: When there is a 4.1, there should be a 4.2 as well. I would suggest to find a subhead for the first paragraphs of this section.

- Figure 4 and related discussion: Is Fig. 4 based on the temporal and zonal mean of the entire data set? Can you say something about the order of the magnitude of the differences, if shorter time periods are considered (monthly means or even shorter).

- P6, L8: climatologies “of what” ?

- P6, L13-15: Is there a reference for the EqL computation used here, so that a reader may be able to look up the computation in detail?

- P6, L17: This is potentially the only greater question which I have. Generally, when you speak of variability, here it is talked about the variability along the polar vortex edge, is this a variability caused by the fact that the reanalyses differ in the representation of the atmospheric features among each other, or because there is a large natural variability of the feature. Maybe it could help to also look at the variability of the shown quantities in individual reanalysis data sets to show whether these already have a large variability or not.

- P7, L7: Do you search the tropopause from top or bottom or asked differently do you refer here to the lowest or highest tropopause in the presence of multiple tropopause as can occur in the vicinity of tropopause folds?

- P7, L29: I wonder whether the parameterizations of orographic GW really have an effect on the tropopause altitude or whether the effect seen here is rather related to larger, resolved GWs themselves above such orography. As far as I know the standard orographic GW parameterizations rather affect higher altitudes by dumping energy at a specified level somewhere in the middle to upper stratosphere and thus affecting the resolved mean flow at those altitudes but not at the tropopause level. Could this here be also a result of the data assimilation, since these are regions with relatively frequent GW occurrences which might be included in radiosonde data which become assimilated?

- In the discussion of Fig. 7 starting on P7, l24, I wonder how much the shown differences could be related to the vertical grid spacing of the individual reanalyses? These data sets all differ in their absolute vertical grid spacing as well as the interpolation may be dependent on the actual location of the individual model levels in the tropopause region. Maybe it would be worth adding information about the vertical grid spacing of the reanalysis products in the tropopause region/stratosphere.

- P9, l4: ...smaller THAN the polar cap

- affiliation 2 and 6 are the same