

Interactive comment on “TPVTrack v1.0: A watershed segmentation and overlap correspondence method for tracking tropopause polar vortices” by Nicholas Szapiro and Steven Cavallo

Anonymous Referee #2

Received and published: 18 October 2018

Reviewer's summary of the manuscript

In "TPVTrack v1.0: A watershed segmentation and overlap correspondence method for tracking tropopause polar vortices," Szapiro and Cavallo present a new software framework for detecting and tracking tropopause polar vortices (TPVs). They qualitatively describe TPVs as persistent areas of high potential vorticity (positive and negative) occurring along the tropopause that are associated with the broader polar vortex. They describe two other TPV tracking methods and give overview of the various considerations involved in robustly detecting and tracking features like TPVs.

C1

The authors summarize their choices for tracking methodology as: (1) a watershed segmentation model on potential vorticity (for negative and positive anomalies separately), combined with (2) an advection-overlap method for ascertaining temporal continuity of detected vortices. The method considers TPVs as detected tracks that persist for 2 or more days and that occur poleward of 60 degrees. They describe a number of geometric and dynamic metrics that TPVTrack v1.0 can calculate.

Szapiro and Cavallo apply TPVTrack v1.0 to an idealized spatial field of potential vorticity—with added noise—to compare the method with two other tracking methods and to demonstrate that in principal TPVTrack v1.0 identifies vortices in a way consistent with the authors's descriptions of the tool. They further apply the tool to a specific synoptic case that involves simulations with WRF at multiple resolutions. Finally, they apply the tool to one year's worth of ERA-Interim output and examine vortex-centered composites of TPVs to show that their detected TPVs fit dynamical expectations.

Summary of Review

The manuscript presents a well-written and thorough description of TPVTrack v1.0, which appears to be a valuable new tool for researchers desiring to track tropopause polar vortices. The methodology described is sufficiently novel compared to other methods, and most of the methodological choices seem logically sound. With a few notable exceptions (described below), the figures are well-described and support the manuscript text. Overall, the manuscript is nominally worth publication in some form.

However, despite this, I have a few major concerns that in combination may preclude publication in GMD in its current form:

* lack of compelling scientific application of the new tool, * superficial discussion of the reasoning behind some key methodological choices, and * superficial discussion of uncertainties and assumptions involved in the methodological choices. * inadequate discussion about how the results of the method might (or actually do) depend on technical details of the model used (e.g., especially horizontal resolution)

C2

Overall, these concerns combine such that it seems to me that this paper is not appropriate for the aims and scope of GMD. Relatedly, I am not sure it would be of general interest to the GMD readership, since the intended audience appears to be solely dynamicists with interests in tracking TPVs, which I expect is a narrow portion of the readership.

The manuscript comes across to me as a sound technical description of a new software tool. While GMD does support manuscripts that offer technical descriptions, these—as far as I am aware—are technical descriptions of new geophysical models. Instead, this manuscript provides technical description of a tool for model analysis. It is not clear to me that this type of manuscript is appropriate for GMD. It would however be a very obvious candidate for publication in a journal like The Journal of Open Source Software.

Even if GMD does support publication of this category of article, I would still suggest that the manuscript warrants major revisions due to the reasons enumerated above. In its current form, the manuscript lacks a compelling scientific application that would aid readers in seeing value and relevance of this method, and the lack of discussion about methodological choices effectively makes for results that are not repeatable by others. For example, if a reader decided to implement this TPV tracking method and had to make a choice about some specific implementation details ('TPVs are defined as tracks with a core at genesis north of 60N lasting at least 2 days' – what was the basis for the authors deciding on this definition of tracks???), the manuscript does not provide sufficient detail for the reader to come to the same conclusion as the authors about the implementation details given in the manuscript. This also then precludes a reader from debating the authors's conclusion about the choice of these details, since details about the choice are not given. Related to this, some of these implementation choices have some amount of uncertainty associated with them (e.g., why 60N and 2 days versus 61N and 1.5 days?), but the authors do not discuss the implications of these uncertainties. This is critical, since there is a growing recognition in the litera-

C3

ture that such details may have enormous and important uncertainties associated with them (e.g., see two recent papers on the Atmospheric River Tracking Method Intercomparison Project: <https://journals.ametsoc.org/doi/abs/10.1175/BAMS-D-18-0200.1> and <https://www.geosci-model-dev.net/11/2455/2018/gmd-11-2455-2018.html>).

Additional details of these criticisms follow.

Major issues

No Scientific application

The manuscript is completely focused on a technical description of the TPV tracking tool. Though GMD does support technical descriptions of models, my experience as a reader is that technical papers are much more useful if they also provide a simple scientific use case that illustrates the value of the new method/model. Without this, the scientific value of this particular TPV tracking method is not clear.

On pg 12, line 16, the authors state "that a TPV climatology paper is beyond the scope of this paper." My initial reaction to seeing this is 'that is a shame'; it seems like it would be a very easy target for the authors to ask the simple questions of "what is the climatology of TPVs?", "how does the climatology from our method compare with that from the other methods that exist in the literature?", and "if there is a difference, why might this method be more valid?" The authors even state on page 11 that they ran their algorithm on ERA-interim from 1979 to 2015: so why not show any of these results? In my opinion, arguing that this is beyond the scope of the paper harms rather than helps the paper.

Note that the authors do seem to have an interesting result associated with Figure 8, but they only devote one sentence to that figure. If I understand the result correctly (but see my comment in the 'Minor Issues' section below, noting that I'm not convinced I do), this implies that TPVs grow, reach a maximum size some time in the middle of their life cycle, and then shrink again. This seems interesting and potentially worth digging

C4

in to a bit more. Is this expected? Would other tracking methods show a similar result?

Lack of Justification for Methodological Choices

In general, the authors do a good job of describing their reasoning behind key methodological choices: e.g., the watershed basin method is more robust to 'grid scale undulations' in the field. However, there are a four key choices for which inadequate justification is provided:

1. 'TPVs are defined as tracks with a core at genesis north of 60N lasting at least 2 days' (pg 11, line 7) * TPV minimum latitude: 60N * TPV minimum duration: 2 days
2. 'Default settings are a 300 km filtering disk for regional extrema and 5th percentile of the amplitudes of the basin's boundary cells with respect to the core' (pg 8, lines 16-17) * TPV filtering disk radius: 300 km * TPV percentile threshold: 5th percentile

Of these four choices, the TPV minimum latitude and TPV minimum duration are the least justified: as far as I can tell, the authors simply state this choice without qualification. If another researcher implemented this method, the paper provides no information about why the researcher should conclude that 60N and 2 days should be the default values. Also, these choices effectively embed assumptions about the nature of TPVs in them, and these assumptions (and their implications) should be explicitly stated.

The TPV filtering disk radius and TPV percentile threshold do have some explanation provided; the authors appropriately explain that 'Increasing the radius for regional extrema generates larger basins and fewer objects. Increasing the restriction percentile will generate larger basins' (pg 8, lines 18-19). However, the authors also state that 'The default settings best match manual tracks in a small set of case studies.' This seems like a reasonable basis on which to make parameter choices, but what small case studies and which manual tracks? Without this information, a reader has no chance of evaluating whether they agree that the given choices do result in a good match with case studies and that another choice of parameters would be inferior. I assume that the authors are partly forward referencing the result stated on Pg 12, lines

C5

9-10: "TPVTrack's track exactly matches our manual track," and if so, this should be made explicit. However, this is only one track (the authors indicate that more than one track is used for deciding on parameters), and the track is not actually shown: these manual track are critical data on which the authors are making methodological choices and so should be included in the paper (perhaps as supplementary material?).

No Discussion of Uncertainties Associated with Methodological Choices

Related to the above, the authors do not adequately discuss the implications of uncertainties associated with these parameter choices. As noted above, the authors do discuss the effects of varying the filtering disk radius and percentile threshold on individual fields. This is a good direction for discussion, and it would be useful if the authors could expand this discussion to (1) include discussion of implications for changing other parameters, and (2) expand the discussion further to include implications for climate-length studies. It would be even more useful if the authors directly showed the effects of these parameter choices on climatological TPV track information.

This is particularly important as there is a growing recognition in the literature that this uncertainty can have major implications for our understanding of weather and climate. For example, the IMILAST project (extratropical cyclone detection intercomparison) shows a ~6x variation in the counts of cyclones across 15 different methods: <https://doi.org/10.1175/BAMS-D-11-00154.1>. Likewise, ARTMIP (atmospheric river detection intercomparison) a similarly large spread in AR statistics across numerous detection methods (<https://journals.ametsoc.org/doi/abs/10.1175/BAMS-D-18-0200.1> and <https://www.geosci-model-dev.net/11/2455/2018/gmd-11-2455-2018.html>). The ARTMIP project is currently working on experiments to understand whether these different algorithms might produce different climate sensitivities for ARs in climate change experiments. Given this growing understanding in the literature, it is critically important that tracking-method papers such as this explicitly explore uncertainty at the outset. I'm not arguing that the authors should tackle an intercomparison of the scale of IMILAST or ARTMIP, but since, as the authors note on pg 8 lines 21-22, TPVTrack makes it easy

C6

to explore parametric uncertainty, the authors should do just that.

Minor issues

pg 1, lines 24-25: "Diagnostic trajectories and prognosed scalar transport further support the advection-dominated dynamics for individual cases (not shown)." <- It is very odd to include a new, not-shown result in the intro: why do this?

pg 2, lines 4-6: This section should reference IMILAST and ARTMIP, which are both very relevant to the discussion

pg 2, lines 18-20: Regarding the first sentence of this paragraph: from where does this qualitative definition originate? If there is a common source (e.g., a textbook), it should be cited. If not, would other polar dynamicists agree on this? Marty Ralph had to convene two AGU townhalls to come to a qualitative, consensus definition of ARs (which is now in the AMS glossary): why would TPVs be different? If this definition is original, I would suggest a rephrasing to make clear that this is a proposed definition: e.g., "We propose a functional, qualitative definition of TPVs: ..."

pg 2, lines 20-21: "...are fundamental to an automated scheme" <-I would argue this is true for any objective, quantitative scheme: whether automated or not.

pg 4, line 6: "through a modular, object-oriented approach is publicly available" <- There seems to be a word missing in this sentence (should it be "*which* is publicly available"?)

pg 8, line 8: 'It is not clear how "optimal" settings would be defined or justified. The default settings best match manual tracks in a small set of case studies.' <- These two sentences seem to contradict each other. The first says we don't know how to define 'optimal', and the second says that we used a small case study to show that our parameter setting results in the best match (which sounds 'optimal' to me).

pg 8, line 30: "...; metrics is independent" ('is' should be 'are')

C7

pg 11, line 25: "similar to values found by trapping 2 PVU by searching down from the model top for these grid scales" <- I have no idea what this means. I would suggest rephrasing somehow.

pg 11, line 30: "(Fig 6.e,h,i)" <-Is the lettering here what was actually intended? I'm having a hard time understanding what the authors are referring to.

pg 12, lines 9-10: "TPVTrack's track exactly matches our manual track." <-What manual track? I see no figure for this.

pg 12, lines 18-19: "Both cyclonic and anticyclonic tracked TPVs reach their minimum radius at the beginning or ends of tracks in the majority of cases (Fig. 8)." I struggled to see how Figure 8 indicates this. It's not that I doubt the result, but rather that the caption for Figure 8 doesn't make sense to me and/or the axis labels are confusing.

pg 13, line 25: "...and the bottom of the stratosphere may reach the surface" <- What!? Perhaps there is a polar atmospheric phenomenon that I've not yet learned about, but I've never heard of the tropopause reaching the surface in any dynamical circumstance. I'm wondering if the wording here conveyed something that the authors didn't intend. Otherwise, if this can actually happen, a reference here should be added, since I expect I wouldn't be the only reader to be surprised to learn this.

Figure 2b: I read the text and caption several times and I still can't figure out what Figure 2b is supposed to convey.

Figure 6: Titles/labels on the subplots would be extremely useful. Given that there are 3 resolutions, my initial inclination was to think that columns correspond to resolution—but this isn't true (d,e,f). Because of this confusion, I found I had to repeatedly keep looking between the figures and the caption to understand what I was looking at. It doesn't help that the captions for the subplots reference other parts of the caption ("(i) as in e, but for..."). I found I spent way more time going back and forth between the caption and figures than I normally do in a paper, which made this quite frustrating—and

C8

I don't think it needs to be.

Figure 8: I think the caption needs to be reworded. I went back and forth between the figure and the text multiple times before I think I understood the figure. If I understand it correctly, it might be more usefully worded as "Probability of TPVs being at their minimum equivalent radius as a function of lifecycle for cyclonic...". Also, for the label of the horizontal axis, I would suggest the word 'lifecycle' rather than 'lifetime', because the term lifetime made me think that the horizontal axis referred to a measure of the duration of the TPV relative to other TPVs.

Interactive comment on Geosci. Model Dev. Discuss., <https://doi.org/10.5194/gmd-2018-180>, 2018.