We thank the reviewers for the constructive comments towards improving the work. The pages below include a point-by-point reply to the reviewers' comments (with our text in blue) and a marked-up manuscript showing the changes using latexdiff.

Sincerely,

Nicholas Szapiro and Steven Cavallo

In response to RC1:

We thank Baird Langenbrunner for review and improvement of the software package and documentation. In addition to his modifications incorporated via pull request, we include additional software changes for the unified branch, described below in response to the review point-by-point. The manuscript has been changed to acknowledge the review.

More documentation: Docstrings below function definitions now include a brief description of what the function does and the input arguments if possibly ambiguous.

Code portability: Migration to Python 3, cartopy, and a Python package are welcome suggestions for development and incorporation into future versions.

Clear list of output: An output section has been added to the User's Guide, describing output from demo() and demo_algo_plots(). The other options at the bottom of driver.py were products of older versions. They have been removed, with driver.py correspondingly cleaned.

"Note that .pyc files will be created (but ignored in .gitignore).": The sentence has been added to the User's Guide in the ERA-Interim test case section.

Documentation contents: With more comprehensive documentation a goal of future development (as a Python package), description of the core modules in the manuscript and User's Guide is intended to further orient the reader. We direct users to the (readable) source code if the desired implementation details extend beyond what is described in the documentation. NetCDF files output from preProcess, basinMetrics, and tracks now include units and long_name metadata. The segmentation and correspondence files contain descriptions metadata.

"info" variable: Changing the "info" variable is now a step in the ERA-Interim test case in the User's Guide.

Reconstruct Anaconda environment: A docs/environment.yml is now included and mentioned in the User's Guide.

test-tpvTrack directory: The test-tpvTrack directory with output from the example test case was added with the reviewer's pull request. Files have been updated to match output from the current version.

Changes in my_settings.py: The User's Guide now correctly instructs the user to modify variables in my_settings.py.

Color map: The colormap was changed with the reviewer's pull request.

Response to Anonymous Referee #2

We thank the reviewer for improving the clarity, content, and quality of the manuscript. Point-by-point replies are included inline below, with the reviewer's text in black and our responses in blue. Note that Figs. 8 and 11 have been added, so the previous Figs. 8 and 9 are now Figs. 9 and 10, respectively.

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.

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)

We respond to these 4 major concerns in the corresponding, more-detailed sections below.

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.

We agree that dynamicists with interests in tracking TPVs are an intended audience of this manuscript. We disagree that they are the sole audience.

TPVs are a class of upper level potential vorticity anomalies usually present in the Arctic. Upper level and surface PV anomalies are classically important in mid-latitude synoptic meteorology. With further connections to and from TPVs throughout the Earth system possible (as outlined in the paper's introduction), additional recent work connects TPVs with cold air outbreaks (Biernat 2017). Connections with atmospheric rivers may be possible as well but have not yet been pursued to our knowledge.

We expect that furthering our understanding of these connections will be fruitful. Publication of TPVTrack in GMD facilitates these goals. Moreover, TPVTrack can be adapted to track other features as mentioned in the Conclusions (with modification done in Biernat 2017 for synoptic cold pools, for example), potentially of additional interest to the GMD 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.

As the reviewer does not further define geophysical models and model analysis or differentiate between them, we consider two definitions. Under the current AMS Glossary's definition

(http://glossary.ametsoc.org/wiki/Model), TPVTrack fits in the “heuristic method” category as a pattern recognition technique as an application/adaptation of image processing strategies. More broadly, if a geophysical model is a scientific model that takes geophysical input and generates geophysical output, where "[s]cientific modelling is a scientific activity, the aim of which is to make a particular part or feature of the world easier to understand, define, quantify, visualize, or simulate by referencing it to existing and usually commonly accepted knowledge" (https://en.wikipedia.org/wiki/Scientific_modelling), TPVTrack qualifies as well. In short, TPVTrack takes input of atmospheric fields, defines physical features termed tropopause polar vortices through rule-based patterns, and outputs a history of features to be post-processed by the user. It is a component towards furthering our understanding of TPVs and their interactions with the Earth system. That is, TPVTrack is a geophysical model.

However, if it is still argued that TPVTrack does not qualify as a geophysical model but rather "a tool for model analysis" instead, we believe that this manuscript would still be suitable for publication under GMD's Method for assessment of models manuscript type as both a software tool and discussion of a novel method for data analysis (https://www.geoscientific-model-development.net/about/manuscript_types.html#item3). In this alternate view, TPVTrack is a diagnostic tool to add TPV-related state variables to the model outputs, where the output data can then also be used independently.

While we are not familiar with the distribution of interests of GMD readership, “CycloTrack” (as a model description) and “TempestExtremes” (as a method for assessment of models) are similar works focusing on surface cyclones that have been published in GMD.

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.

We respond in the major issues section below. Note that, using the included TPVTrack software and available input data, all methods and results are reproducible. We invite competing or

alternate conclusions regarding the formulation, implementation, and internal settings of TPVTrack.

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 literature 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).

We respond in the major issues section below.

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.

An additional use case has been added to the introduction. In the Conclusions, scientific value is now better communicated by presenting (1) selected results from the default settings and (2) concerns over use of a single approach.

This manuscript is focused technically in order to document the new method. The manuscript also stresses that a unique method may not be the best strategy for accurate representation of TPVs in all respects and complementary approaches may be fruitful. Originally, TPVTrack was initiated to address limitations in H05 in terms of robustness of spatial shape to small scale noise significant to local gradients (which is particularly important for higher-resolution datasets) and 1-1 temporal history (which is particularly important for TPVs, which are UTLS

features that can “live” for months). These problems also plague other existing approaches. Exploration of the physical basis versus artifact character of existing approaches led to TPVTrack, and we hope the future holds further progress as well.

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.

We agree that an updated TPV climatology is worthwhile and have added “Note that a TPV climatology is beyond the scope of this paper and the focus of a separate work.” This publication is one of a series of papers from members of the Arctic and Antarctic Research Group at the University of Oklahoma. Another graduate student is leading paper(s) on such a climatology, with discussion of input reanalysis, seasonality, associations with teleconnections, long-term trends, significance of differences, and sensitivity to tracking method. It is not clear to us that separating the climatology as a separate work harms this paper, and the separation benefits the climatology through fuller treatment.

Note that Sect. 3.2 Historical test cases is based on tracks using ERA-Interim input data from 1979 to 2015. The vortex-centered composites consider only one year of the historical period to reduce the cost of re-centering about each TPV.

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

Strictly, now Fig. 9 shows that the minimum radius during a TPV’s life tends to occur during the beginning or end of the TPV. The reviewer’s implied lifecycle is consistent with this pattern. However, so are TPVs that monotonically grow or decay over life. An analogous figure for the normalized lifetime of maximum radius does show that the maximum radius largely occurs between genesis and lysis. This is now noted but not shown.

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.

These are the four user-defined parameters in TPVTrack. We reiterate that these are additional degrees of freedom that may impact analyses associated with TPV tracks. Rather than guarantee one set of fixed values, sensitivities should be explored by a user as for any model. We cannot claim that the settings are universally optimal, but they are reasonable for tested and anticipated cases. Rationale for bounds for the parameters have been added to Sect. 2.2.6 Parameter Settings.

Several tradeoffs of the size of the filtering disk and restriction percentile are discussed in the text and illustrated concretely in Figs. 6 and 7. To subset tracks into TPVs associated with the polar vortex that last, 60N is the mean latitude of the polar jet and 2 days matches the definition in H05. These have been added to Sect. 3.2, where the criteria are 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 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?).

The four periods of interest used to inform parameter choices have been listed in Sect. 2.2.6 Parameter Settings.

Since there are dozens of tracks at a given time, comprehensive presentation of all manual maps and TPVTrack tracks (including location and shape) is not simple. Nor is it necessary for the reader's understanding in our opinion as the discussion of the 2006 long-lived track largely covers the pertinent points. Note that the user can easily explore any period of interest by adapting the example ERA-Interim test case and examining the output plots.

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

As the reviewer notes, sensitivities to filtering disk radius and percentile threshold are discussed (in combination with varying input data) for the summer 2006 WRF case. To address points (1) and (2), discussion and illustration of the impacts of latitude and lifetime criteria on mean TPV density have been added (end of Sect. 3.2.2 and Fig. 11).

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?

We believe that it is important to have context for the expected dynamics of TPVs before considering their tracking. A supporting result comparing the track of the long-lived 2006 TPV with a trajectory model of the core is included in Sect 3.2.1 and referenced in the introduction.

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

Reference to ARTMIP has been added. IMILAST was already referenced.

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: ..."

The sentence in question states that “Synoptically, TPVs are coherent anomalies on the dynamic tropopause associated with the larger polar vortex that have a regional minimum in potential temperature and cyclonic circulation or regional maximum in potential temperature and anticyclonic circulation that last over time.”

In the current AMS glossary’s definition of “Polar Vortex” (http://glossary.ametsoc.org/wiki/Polar_vortex; in reference to Cavallo and Hakim 2010), “The term “polar vortex” is sometimes used in reference to smaller-scale (meso- to synoptic scale) vortices that usually occur within the tropospheric polar vortex in polar regions near the

tropopause—for example, “tropopause polar vortices.”” Since the manuscript’s text is different than the AMS Glossary’s partial definition, we have added the suggested text to clarify that this is a proposed definition.

We do not believe that polar dynamicists in consensus would disagree that a planetary polar vortex exists, TPVs are associated and of smaller scale, and TPVs are coherent over time with (nearly-)balanced temperature and wind structure. Of course, the boundaries of a given vortex can vary by definition and possibly should vary by application, as noted in the Conclusions that the “size of a TPV may refer to a number of scales…”.

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.

“automated” has been changed to “rule-based” throughout

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"?)

The sentence is long but grammatically correct: “An implementation…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).

Rephrased “optimal” to “universally optimal” to clarify that the best settings for all applications are unclear. The case studies have also been listed, which further clarifies the lack of universality.

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

The modules have been rephrased for grammatical parallelism.

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.

The sentence has been rephrased and now also references Sect. 3.2.3 on diagnosis of the tropopause.

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.

Added “respectively” to the text. Together with the subplot labels in a later comment, we believe the construction is clearer.

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

Added Fig. 8 showing TPVTrack’s track and a map of one time as used for manual tracking.

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.

The caption has been rephrased.

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.

Reference to a published figure of a cross-section through a PV tower with the tropopause at the surface has been added.

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

The caption has been rephrased. Arrows like in 2a have been added for 2b.

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 I don’t think it needs to be.

The subplots have been labelled with the corresponding grid spacing, closed contour percentile, and filter radius for regional extrema.

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 refereed to a measure of the duration of the TPV relative to other TPVs.

The caption has been rephrased. Given uncertainty in the stage(s) of a TPV’s lifecycle between genesis and lysis, we prefer the term normalized lifetime over lifecycle. This also follows other usage in the literature (e.g., Kew et al. 2010, Fig. 10).

Sincerely,
Nicholas Szapiro and Steven Cavallo

References:

Biernat, K.: Linkages Between Tropopause Polar Vortices and the Development of Cold Air Outbreaks Over Central and Eastern North America, Thesis (M.S.)--State University of New York at Albany, 2017.; Publication Number: AAT 10683411; ISBN: 9780355505092; Source: Masters Abstracts International, Volume: 57-01.; 119 p.

Flaounas, E., Kotroni, V., Lagouvardos, K., and Flaounas, I.: CycloTRACK (v1.0) – tracking winter extratropical cyclones based on relative vorticity: sensitivity to data filtering and other relevant parameters, Geosci. Model Dev., 7, 1841-1853, https://doi.org/10.5194/gmd-7-1841-2014, 2014.

Ullrich, P. A. and Zarzycki, C. M.: TempestExtremes: a framework for scale-insensitive pointwise feature tracking on unstructured grids, Geosci. Model Dev., 10, 1069-1090, https://doi.org/10.5194/gmd-10-1069-2017, 2017.

TPVTrack v1.0: A watershed segmentation and overlap correspondence method for tracking tropopause polar vortices

Nicholas Szapiro[1] and Steven Cavallo[1]

[1]School of Meteorology, University of Oklahoma, Norman, OK, USA

Correspondence: Nicholas Szapiro (nick.szapiro@ou.edu)

Abstract. A new modular, free software package is described for tracking tropopause polar vortices (TPVs) natively on structured or unstructured grids. Motivated by limitations in spatial characterization and time tracking within existing approaches, TPVTrack ~~leverages the~~ mimics the expected dynamics of TPVs to represent their (1) spatial structure, with variable shapes and intensities and (2) time evolution, with mergers and splits. TPVs are segmented from the gridded flow field into spatial objects as restricted regional watershed basins on the tropopause, described by geometric metrics, associated over time by overlap similarity into major and minor correspondences, and tracked along major correspondences. Simplified segmentation and correspondence test cases illustrate some of the appeal, sensitivities, and limitations of TPVTrack, including effective representation of spatial shape and reduced false positive associations in time. Tracked TPVs in more realistic historical conditions are consistent in bulk with expectations of life cycle and mean structure. Individual tracks are less reliable when discriminating between multiple overlaps. Modifications to track other physical features are possible, with each application requiring evaluation.

1 Introduction

Among the disturbances on the extratropical tropopause, tropopause polar vortices (TPVs) are common, coherent upper-level potential vorticity anomalies with typical radii of 100 to 1000 km and lifetimes of days to months (Hakim, 2000; Hakim and Canavan, 2005). TPV-like features play documented roles in so-called potential vorticity thinking (e.g., Hoskins et al. 1985; Holton and Hakim 2013), surface cyclogenesis and high-impact weather (e.g., Davis and Emanuel 1991; Morgan and Nielsen-Gammon 1998; Grams et al. 2011; Simmonds and Rudeva 2012), geographic climate (e.g., Shapiro et al. 1987; Nieto et al. 2008; Kew et al. 2010), stratosphere-troposphere exchange (e.g., Sprenger et al. 2007), potential predictability (e.g., Provenzale 1999), and evaluation of model skill (e.g., Béguin et al. 2012). Knowledge of ~~the structure and history of TPVs~~ where a TPV is, how it is shaped and changing, and what it is near may aid further systematic and reproducible approaches in these and more areas. For example, TPV tracks can aid a mechanistic understanding of how information from distant radiosonde observations can impact the development of a surface cyclone days later (Yamazaki et al., 2015). Additionally, can an individual long-lived TPV contribute a significant circulation on seasonal scales? Accurate, automated tracking of TPVs in gridded data improves our knowledge of TPV structure and history. Within a framework of Ertel's potential vorticity (Rossby, 1939; Hoskins et al., 1985; Pedlosky, 2013) in a composite sense (Cavallo and Hakim, 2010), dynamics of TPVs are dominated by quasi-horizontal

advection subject to generally smaller diabatic and frictional forcings. Diagnostic trajectories and prognosed scalar transport further support the advection-dominated dynamics for individual cases (~~not shown~~e.g., Sect. 3.2.1).

Tracking approaches for various features follow an approach of spatial identification and time correspondence ~~(Hodges, 1999; Limbach e~~ (Hodges, 1999; Limbach et al., 2012; Neu et al., 2013; Ullrich and Zarzycki, 2017; Shields et al., 2018). However, choices and details within and between approaches can lead to systematic differences and uncertainties in results, both from algorithm sensitivities and the dynamics of the feature. For tropical cyclones, sensitivities to thresholds for size, wind speeds, warm cores, genesis, and duration impact track statistics (Walsh et al., 2007; Horn et al., 2014). For extratropical surface cyclones, deeper and stronger features are more consistent and robust across methods (Neu et al., 2013; Tilinina et al., 2013). Differences in tracks across algorithms, especially for features without independent or consensus definitions like TPVs, complicate use and interpretation as each approach is perfect within its own circular definition (Neu et al., 2013). Because our primary interest is in the physical TPVs rather than the tracked objects, clarity in the characteristics of each algorithm may lend confidence and rationale for informing whether results derived from automated tracks are artifacts or physical, especially if an ensemble with complementary approaches can be constructed by opportunity (e.g., Tebaldi and Knutti (2007)) or design.

Towards such a usable description of TPVTrack, Sect. 2 frames TPV tracking approaches and describes the current algorithm design and implementation. Idealized segmentation and correspondence test cases with reference truths quantify some of the appeal, sensitivities, and limitations of components of TPVTrack, including relative to several existing approaches (Sect. 3.1). Individual and aggregated historical cases relate the tracked objects to physical expectations of TPVs (Sect. 3.2). Section 4 concludes with a summary and possible extensions to other features.

2 TPV Tracking

Synoptically, we propose a functional, qualitative definition that TPVs are coherent anomalies on the dynamic tropopause associated with the larger polar vortex that have a regional minimum in potential temperature and cyclonic circulation or regional maximum in potential temperature and anticyclonic circulation that last over time. Defining “coherent,” “regional,” “associated," and “last” are fundamental to ~~an automated~~ a rule-based scheme. Note that the existence of a local extremum in a continuous surface implies the existence of a closed contour about that extremum, thus consistent with a notion of vortices (rather than waves) with trapped fluid for frictionless and adiabatic flow (Hakim and Canavan, 2005).

2.1 Existing algorithms

We summarize the automated approaches we know have been used to track and characterize TPVs or similar features in terms of multiple steps: choice and pre-processing of gridded data, segmentation in space, characterization of features, and tracking in time. Several limitations of existing approaches motivated the design and implementation of TPVTrack. With further details in the references, here we focus on comparing the interdependent choices that define “regional,” “coherent,” and “lasts.” Defining “associated” with the larger polar vortex is typically a post-processing step and applicable to any set of tracks, e.g., Hakim and Canavan (2005) subset an “Arctic" category of tracks that last at least 2 days with over 60% of their lifetimes north of 65° N.

Each method bases objects about extrema. Since a surface can have many local extrema, to target salient features, filtering criteria are imposed so not all local extrema are cores of objects; an extremum must be "regional." For Hakim and Canavan (2005), using 2.5° data, any local extremum must also be an extremum within a 650 km radius. For Kew et al. (2010), background potential vorticity is calculated by averaging over a large (10^7 km^2) area. Positive anomalies are candidate extrema, and a flow-dependent contour of smaller prescribed area (1.7×10^5 km^2) is computed about each candidate. Lesser maxima within the bounding contour of a stronger maximum are disregarded. Simmonds and Rudeva (2014) require a local extremum in the surface Laplacian and minimum concavity for saliency through a threshold (unspecified) on an area-averaged Laplacian. In effect, choices for saliency and neighborhoods are imposed filters, integral to each approach and with a number of trade-offs as is typical of filter design.

What area is associated or "coherent" with a regional extremum? Spatially, where is the edge of the feature? For Hakim and Canavan (2005), eight spokes are extended from the vortex core until the radial gradient in potential temperature changes sign for each. The minimum of these potential temperature values across all eight radials defines the bounding isentrope. For Kew et al. (2010), the flow-dependent contour of prescribed area also defines the feature in space. If overlap occurs between two features, the weaker anomaly shrinks to a smaller contour yielding no overlap. For Simmonds and Rudeva (2014), radius is the weighted mean distance from the core to where the Laplacian vanishes. Consider how these definitions would represent two TPVs, one physically larger than the other. For Kew et al. (2010), the area is prescribed and the same for both TPVs. For the local gradient-based methods, sensitivity to smaller-scale noise in the surface can trigger stopping criteria with resulting areas somewhat arbitrary. Note that for some approaches (e.g., point methods) spatial shape is simply an additional diagnostic, whereas other approaches may leverage the additional information (e.g., for overlap tracking).

When is a TPV the same feature as at the previous time? How long does an individual TPV last? Hakim and Canavan (2005) use a simple proximity algorithm where a track is extended if a vortex core is within 600 km of the previous location (advected previous location for later versions). Corresponding to a maximum vortex speed of 28 m s^{-1}, the threshold distance subjectively works for the intended polar region and also is easily reproduced. Kew et al. (2010) develop an approach akin to contour surgery, where cells are advected forward in time by analyzed winds, and the resulting shape is from a convex hull of the destinations after exclusion of outlier locations. Significant (10%) spatial overlap extends a track. Note that both of these extension strategies duplicate the entire history of a TPV if there are multiple matches. Simmonds and Rudeva (2014) estimate a new position of each cyclone by weighting the track's past and climatological velocities, calculate probability of associations based on distance and central pressure differences, and keep the combination of associations within candidate groups (constructed by linking all positions within a cutoff distance) that maximize the product of pairwise probabilities as tracks. As reflected in the summary above, associations in time are decisions for whether the expected evolution sufficiently matches the actual states. Defining an expected evolution can involve a range of assumptions and complexity impacting fidelity, but approaches may be related notionally. For example, using a prescribed search radius between point locations is equivalent to calculating horizontal overlap for features with prescribed shapes of disks and no transport. Comparisons are illustrated more concretely in Sect. 3.1.

2.2 Description of TPVTrack

An implementation in Python 2.7 of this algorithm supporting and tested with output from the National Center for Environmental Prediction's Global Forecast System (GFS; e.g., Saha et al. (2010)), European Center for Medium Range Weather Forecasting interim reanalysis (ERA-Interim; Dee et al. (2011)), Weather Research and Forecasting Model (WRF; Skamarock et al. (2005)), and atmospheric component of the Model for Prediction across Scales (MPAS-A; Skamarock et al. (2012)) through a modular, object-oriented approach is publicly available (Sect. 4). A user's guide is included with the package. The following outlines the core modules in TPVTrack. Each module outputs one NetCDF (Rew and Davis, 1990; Whitaker, 2015) file.

2.2.1 Input data

The meteorological inputs to the algorithm are gridded zonal wind, meridional wind, relative vertical vorticity, and potential temperature on the extratropical tropopause (Gettelman et al., 2011; Ivanova, 2013) over time. With no diabatic or frictional forcings, an air parcel would have fixed potential vorticity and potential temperature. With small forcings for the associated parcels, TPVs can be tracked as material eddies on the dynamic tropopause (with preferred diagnosis described in Sect. 3.2.3).

A subset of the original domain ($[-\frac{\pi}{2}, \frac{\pi}{2}] \times [0, 2\pi)$ in radians latitude/longitude) may be specified as the domain of interest. Within the domain, values are interpreted in a finite volume sense as areal means located at cell centers. In deriving the tropopause surface, some data sources have missing values (e.g., GFS uses a missing value if the entire column is above the potential vorticity threshold). If no additional information is available, we fill missing values in pre-processing by an iterative extrapolation from valid nearest neighbors. We do not track globally since a potential vorticity isosurface does not correspond to the tropopause near the tropics (e.g., Highwood and Hoskins (1998)). The default is to track polewards of a user-specified latitude, set to $\pm 30^\circ$ for the northern and southern hemispheres, respectively, as the average position of the subtropical jet stream to be inclusive of candidate TPVs. Especially since we are interested in polar regions, we treat the longitudinally duplicated north and south poles in latitude/longitude grids specially to be consistent with the finite volume interpretation throughout the rest of the grid. Each physical pole is treated as a single polar cap with the pole neighboring number of longitude points .

To handle various data sources modularly, each is supported natively by Mesh and Cell classes defining the geometry and topology of the domain. The Cell class provides methods for obtaining the location, area, nearest neighboring cells, and cells within a specified radius of a cell in the mesh. For example, the left neighbor of a cell in a latitude/longitude mesh with index (iLat, iLon) is (iLat, (iLon-1) mod nLon). The Mesh class provides methods for the latitudes, longitudes, and areas of all cells and finding the closest cell to a point. This allows for optimizations within a unified code implementation. For example, consider finding the cell center closest to a point in a latitude/longitude grid versus an MPAS Voronoi unstructured mesh. For a latitude/longitude grid, we can leverage the underlying structure to simply find the closest latitude and longitude independently. For an unstructured mesh, the provided connectivity of cell-to-cell neighbors can be used. Starting from a guess cell, the closest cell to the point can be found by iteratively walking towards a closer neighbor until no closer neighbor is

found. Both approaches involve fewer operations than a brute force global minimum of all cells' distances to the point. Various meshes could be unified through a common background data structure (e.g., spatial quadtree), but this is not required.

2.2.2 Spatial segmentation

Segmentation maps the mesh cells to disjoint spatial objects at each time (Fig. 1). Model tropopause data contain signals on a spectrum of wavelengths, from planetary to grid scales. These scales come from both the originating model and the post-processing that interpolates to the tropopause. As illustrated in Sect. 3.1.1, robustly segmenting TPV-scale objects requires some filtering of smaller scales. Two key questions are: what counts as an extremum? What area is associated to an extremum?

A value is a local minimum if none of the values of its nearest neighbors are smaller. Local maxima can be identified by searching for local minima on the negative of the surface. There are currently two options to define a regional minimum within a prescribed region (e.g., 300 km radius disk): the smallest value or the smallest local minimum. While qualifying as the smallest value requires a comparison to all other values in the region, qualifying as the smallest local minimum is a comparison only to other local minima. A minimum across the region is a stronger filter that will result in fewer objects, while only comparing to other minima permits distinct objects as long as the minima are sufficiently separate. This may be beneficial for a tropopause with undulations within a significant larger-scale gradient. The distinction vanishes for a surface with numerous local minima. The prescribed area is a fundamental user-defined filter for a scale of saliency. Alternately, basins with less than a minimum amplitude could be filtered by merging into neighbors. This option has not been explored, since it precludes identifying weak amplitude TPVs, and a reference potential temperature is needed.

For the area coherent with an extremum, local growing methods suffer from a noisy surface triggering stopping criteria (e.g., Sect. 3.1.1), so a broader perspective is necessary to be more robust. Considering the tropopause as a surface with local undulations, watershed basins naturally represent the spatial objects. A watershed basin is defined by a minimum and a basin that drains by steepest descent to the minimum. Particularly given the grid scale undulations, a direct watershed transform of the surface results in a classical over-segmentation (Serra, 1983) as each local minimum forms a separate object. To resolve the over-segmentation produced by small scale noise, we implemented a variation of a marker-controlled watershed transform with the regional minima as markers.

The regional watershed basins are formed in two steps. First, each cell is associated to a local minimum by following a path of local steepest descent through nearest neighbors. In case of a plateau, the cell with the smaller global index is chosen for consensus. While the occurrence of bit-equal values is rare in general, lossy data compression, idealized conditions, and other systematic factors can increase the likelihood of plateaus. Second, to form basins only associated with regional minima, each local minimum that is not a regional minimum is redirected to drain into the most intense minimum in the neighborhood, iterating until each local minimum drains into a regional minimum.

In order to segment a surface into cyclonic and anticyclonic TPVs, we first calculate watershed basins from the potential temperature surface (lows) and the negative of the potential temperature surface (highs), so each cell is initially mapped to two basins. Then, all regional extrema are mapped to themselves, and each non-extremum cell is mapped to its corresponding either high or low basin by the sign of its local relative vorticity (Fig. 1.b). Thus, an anticyclonic non-extremum cell is mapped

to its high regional extremum. Note that – since highs and lows are assigned based on vorticity – objects are disjoint, but each individual basin is not necessarily connected. While connectivity can be imposed by, say, flood-filling from the core of each vortex, doing so can remove filamentary sub-structures possibly worth preserving.

Since the resulting segmentation is driven by steepest descent of the input surface, the segmentation is sensitive to undulations in the foothills away from a vortex core. To reduce this sensitivity and better isolate the anomalies, optionally, a bounding contour can further restrict each basin (Fig. 1.c). The cells outside the bound are mapped to a background. We identify boundary cells within each basin as cells with a neighbor not in the basin. A user-defined percentile (e.g., 10^{th}) of boundary cell potential temperature amplitudes with respect to the core then sets the contour independently for each basin. A last closed contour approach (0^{th} percentile) is noticeably less robust to outliers causing severe over-restriction, especially as the originating mesh spacing decreases (Sect. 3.2.1). Restricting the basins tends to produce smaller, more consistent objects over time. This consistency is appealing both physically and algorithmically to improve subsequent correspondence and tracking.

2.2.3 Shape metrics

The flow field associated with a TPV is expected to be related to the TPV's shape (e.g., Thorpe 1986; Masarik and Schubert 2013). Principally, a given idealized potential vorticity anomaly is partitioned between static stability and vorticity; a broad and shallow feature has more stability, while one that is narrow and deep has more vorticity. Moreover, the spatially larger an anomaly, the stronger the induced circulation (Hoskins et al., 1985). Towards building physical understanding and connections, various properties of a basin's geometry may be quantified by using the segmentation map of cells to objects as masks.

We organize these metrics by dimensionality: point, boundary, and area metrics. Point metrics include the latitude, longitude, and potential temperature of the core and maximum amplitude (maximum minus minimum potential temperature in the basin). Boundary cells are identified as cells with a nearest neighbor not in the same basin. Metrics include circularity (length of the perimeter with respect to the length of the perimeter of a circle with equivalent area), eccentricity (ratio of fit major and minor axes), boundary potential temperature (median of the potential temperatures of boundary cells) and boundary amplitude (difference of the boundary potential temperature from the the potential temperature of the core). Area metrics are area-weighted integrals over the basin, including area, equivalent radius (of a planar disk with equivalent area), mean vorticity, and circulation. Incorporating a vertical dimension to a TPV's shape is not straightforward. A TPV's vertical extent does not extend throughout the entire column. Extending a mask on the tropopause to nearby vertical levels would need to account for the possibility of vertically tilted flow.

2.2.4 Time correspondence

Correspondences associate spatial objects between consecutive times. Motivated by the advection-dominated dynamics, the first step is to identify possible correspondences using similarity measured by "horizontal" plus "vertical" overlap. As depicted in Fig. 2, consider candidate basins A at time t_0 and B at time $t_0 + \Delta t$.

Horizontal overlap is calculated by advecting A forward $\Delta t/2$ and B backward $-\Delta t/2$ in time on a sphere using the local winds at t_0 and $t_0 + \Delta t$, respectively (Fig. 2.a) (the "half-time tracking" advocated for in Hewson and Titley (2010)). We use

a simple first order scheme where each cell center in basin A is advected on a sphere by its wind to

$$\phi(t_0 + \frac{\Delta t}{2}) = \phi(t_0) + \frac{v(t_0)}{R_E}\frac{\Delta t}{2} \tag{1}$$

$$\lambda(t_0 + \frac{\Delta t}{2}) = \lambda(t_0) + \frac{u(t_0)}{R_E \cos(\phi(t_0))}\frac{\Delta t}{2} \tag{2}$$

for latitude/longitude ϕ/λ, zonal (u) and meridional (v) velocities, and earth radius R_E. If the resulting ~~$|\phi(t_0 + \frac{\Delta t}{2})| > 90°$~~ $|\phi(t_0 + \frac{\Delta t}{2})| > \frac{\pi}{2}$ the trajectory has crossed a pole, and we adjust: ~~$\phi(t_0 + \frac{\Delta t}{2}) = 180° \cdot \phi(t_0)/|\phi(t_0)| - \phi(t_0 + \frac{\Delta t}{2}); \lambda(t_0 + \frac{\Delta t}{2}) += 180°$~~ $\phi(t_0 + \frac{\Delta t}{2}) = \pi \cdot \phi(t$ valid for both poles. The advected location tags the closest cell as within the advected feature. B is advected back in time, and overlap occurs in the commonly tagged cells. For vertical overlap at the common time, the extreme potential temperatures of A and B are held fixed, as if adiabatic. The intersection of the potential temperature ranges is the overlap (Fig. 2.b).

Concisely, Equation 3 defines the similarity between basins A and B at the next time:

$$S(A, B) = \frac{A_{\Delta t/2} \cap_{\phi,\lambda} B_{-\Delta t/2}}{\max(A_{\text{area}}, B_{\text{area}})} + \frac{A \cap_{\Theta} B}{\max(A_{\Delta\Theta}, B_{\Delta\Theta})} \tag{3}$$

(where $\Delta\Theta$ is the maximum amplitude) as the sum of intersections in the horizontal and vertical, respectively normalized by the maximum possible overlap so that large areas and amplitudes are most similar to other large areas and amplitudes. Similarity is only considered for basins with horizontal overlap. Note that this also mimics how we create correspondences subjectively, in that we look for features of expected intensities in expected locations. Generically, these correspondences define candidate connections between overlapping basins and form a directed graph. Note that, technically, the t_0, $t_0 + \Delta t$, and common correspondence grids could all be different. Such an option has not been explored but could be beneficial for time-varying or sharply non-uniform meshes.

The type of correspondence is then classified. While TPVs can form, decay, split, merge, and persist over time, we assume that a split or merger event is characterized by one primary and then secondary branches. We categorize the primary and secondary branches as major and minor correspondences, respectively. A major correspondence is a 1-1 correspondence between A and B where both (1) B is the most similar to A of all the basins at time $t_0 + \Delta t$ and (2) A is the most similar to B of all the basins at time t_0. All other correspondences are minor.

Errors in correspondences depend on the imposed similarity function. A basin with small area or amplitude may have less chance of subsequent overlap. When there are multiple TPVs with some overlap, similarity determines major correspondences quantitatively. While we developed various alternate cost functions to define similarity based on persistence of aspects of a basin's shape over time, each basin metric has weaknesses leading to non-robust associations. For example, radius changes occur when TPVs split, merge, or the basin foothills change. Intensity and amplitude can jump from data assimilation increments or strong interactions with surface lows and terrain. Practically, the inclusion of various metrics is crucial, but relative weighting requires tuning with some level of subjectivity. The (physically motivated) overlap similarity function is at least as robust for several case studies during development.

2.2.5 Time tracking

To generate tracks for individual TPVs as material eddies with histories of genesis, maintenance, and lysis, major correspondences are connected over time. A track begins when a basin has no major correspondence to a basin at the previous time and ends when a basin has no major correspondence to a basin at the next time. For example, a TPV breaking off from a parent would be tracked from the timestep after the break through the chain of major correspondences. This approach avoids duplicating the entire history of a track when a TPV splits, with such repeat counting yielding inconsistencies with the actual flow state that can potentially impact subsequent composites and climatologies. To determine if a correspondence at a given time starts a track, we check whether the basin is in a major correspondence at the previous time. Without further information, all major correspondences from the initial time start tracks and to the final time end tracks. The tracks resulting from connecting major correspondences can then be further filtered as "polar" by restricting to tracks poleward of a jet stream, latitude, or other criteria depending on the application.

Here, only the major correspondences are used, in forming tracks. However, the major and minor correspondences together describe a tree structure of potential overlaps of vortex air masses. For example, this tree has been used to identify all TPVs at future times that could have overlapped with specified TPVs at given times.

2.2.6 Parameter settings

~~Default~~ To define TPVs with TPVTrack, there are four user-specified parameters: size of the filtering disk for regional extrema, percentile for watershed basin restriction, minimum lifetime, and polar classification. For the first two parameters, default settings are a 300 km filtering disk for regional extrema and 5^{th} percentile of the amplitudes of the basin's boundary cells with respect to the core for restriction. The scale of the filtering disk should be larger than the grid scale but smaller than a TPV. The percentile should be between $(0, 100)$ for more robust and consistent spatial shape than using the last closed contour or the full watershed basin (Sect. 3.2.1). TPVTrack output does not depend on the latter two parameters; rather, the user imposes them in post-processing TPVs from tracks. Minimum lifetime should be chosen larger than the time interval of the input data but less than the lifetime of a TPV. TPVs are classified to be directly associated with the polar vortex if polewards of the (flow-dependent) polar jet (e.g., at genesis or during their lives).

Fundamentally, these are imposed filters, with sensitivities that can propagate through the algorithm. Increasing the radius for regional extrema generates larger basins and fewer objects. Increasing the restriction percentile will generate larger basins. Larger basins have more chance for some correspondence overlap and so may be preferentially favored to last. It is not clear how "universally optimal" settings would be defined or justified in general. The default settings subjectively best match manual tracks ~~in a small set of case studies~~during several periods of interest (August 2006, July 2007, August 2012, and January 2014) in terms of track and basin shape. User-defined parameters are exposed in one settings file, and it is straightforward to vary them towards assessing sensitivities within a fixed tracking approach.

2.2.7 Computational cost and acceleration

For $0.5° \times 0.5°$ ERA-Interim data from 01 January 2010 00 UTC through 30 June 2015 18 UTC every six hours, TPVTrack run serially takes 25.44 wall-clock hours per tracked year. Of the total run time, pre-processing, segmentation, metrics, correspondence, and tracks take 1.9%, 78.3%, 1.4%, 7.4%, 11.0%, respectively, including input/output. Timings are from an HP Z600 Workstation (eight Intel Xeon Processor E5620s at 2.40 GHz with 11 GB of Cache memory and 16 GB of RAM) with Linux version 2.6.32-504.1.3.el6.x86_64 and Python 2.7.12 from Anaconda.

To reduce wall-clock time, a number of optimizations are possible. A more efficient implementation of a discrete watershed transform leveraging sorting (Roerdink and Meijster, 2000) could be used for segmentation. While the modules are sequential, each has an embarrassingly parallel decomposition as written (with modifications to file input/output). Within the current version, pre-processing the input data is independent over time; segmentation is independent over time; calculation of metrics is independent over time; ~~correspondence~~ categorization of correspondences is independent over basin and neighboring times; construction of tracks is independent over track. Each process can write separate, contiguous output files that are then concatenated into one file to maintain consistency with the serial version. Parallel versions of the segmentation and tracks modules have been implemented using Message Passing Interface in Python (Dalcín et al., 2008) with near-linear speedup for each component and identical results to the serial version (for 4 processes). The only communication between processes is global barriers between components since the modules are sequential and output is written to file.

3 Evaluation of TPVTrack

Does TPVTrack accurately represent and track TPVs? Ultimately, any validation (Roache, 1998) depends on a reference truth. We explore this question through both simplified component test cases with reference solutions (Sec. 3.1) and historical cases with more realistic conditions but less clear expectations (Sec. 3.2). A complete discussion, if possible, is beyond the scope of this paper. Rather, we intend the following to highlight characteristics of the approach and direct future sensitivity analyses and intercomparisons. Users focusing on case studies are encouraged to test robustness of individual tracks, while systematic artifacts may be more relevant for climatologies.

3.1 Idealized test cases

Spatial segmentation and time correspondence are central to all approaches. We construct a simplified test case for each – Sect. 3.1.1 and Sect. 3.1.2, respectively – to better isolate and assess specific differences. TPVTrack is compared to our implementations of Hakim and Canavan (2005) and Kew et al. (2010) as reference point and areal methods, respectively termed H05 and K10.

3.1.1 Spatial segmentation

For segmentation, consider a surface θ (Fig. 3.a) and vertical relative vorticity ζ motivated by the Sander's analytic model (Bluestein, 1992) with

$$\theta(x,y) = \sum_i A_i \cos\frac{2\pi}{L_i}(x+\lambda_i)\cos\frac{2\pi}{L_i}y \tag{4}$$

$$\text{sign}(\zeta) = \text{sign}(\nabla^2\theta) \tag{5}$$

where A=(10,1,0.1) K, L=(1000,100,10) km, and $\lambda = L/4$ mimic various scales of undulations, as for the tropopause (Shapiro et al., 1987). Results are depicted in Fig. 3 on a grid with 50 km spacing.

While the 8 main regional lows are represented in all approaches, the shapes of the objects reflect differences in the algorithms. For H05, the gradient radially outward from the extremum first changing sign defines the bounding contour. Local noise in the surface is amplified by gradients and can drastically shrink objects (e.g., objects near (x,y)=(300,2000) versus (1400,1000) in Fig. 3.b), which may be mitigated with suitable prior filtering. For K10, the maximum area of each object is prescribed (1.7×10^5 km^2), with smaller objects resulting from overlapping areas (e.g., the two y=100 objects in Fig. 3.c). While the low and high watershed basins are more robust to local undulations (Fig. 3.d,e) because of the regional extrema filter, the assignment of cells to objects is entirely dependent on the sign of the cell's vertical vorticity. Centered finite differences on the raw field amplify small scales not coherent with the expected larger scale features (Fig. 3.f). Reducing the small-scale noise in the vorticity (Fig. 3.g) is more consistent with the larger scale perspective (Fig. 3.h).

There are also differences in the objects along x=0. For H05, the boundary values are not regional minima given the 650 km search radius. For K10, the 0.5 K contour interval we used for growing each object in turn filters the boundary local minima as parts of existing, more intense objects. For the watershed markers, the 300 km search radius yields boundary objects when aligned with the vorticity. In TPVTrack, this filtering radius is user-defined.

3.1.2 Time correspondence

Consider individual disks of radius r and potential temperature range $[\theta_{min}, \theta_{max}]$ advecting about a domain with

$$x(t+\Delta t) = x(t) + u(t)\cdot\Delta t; \quad y(t+\Delta t) = y(t) + v(t)\cdot\Delta t \tag{6}$$

$$u(t+\Delta t) = u(t) + \sigma_u; \quad v(t+\Delta t) = v(t) + \sigma_v \tag{7}$$

$$\theta_{max}(t+\Delta t) = \theta_{max}(t) + \sigma_\theta; \quad \theta_{min}(t+\Delta t) = \theta_{min}(t) + \sigma_\theta; \tag{8}$$

$$r(t+\Delta t) = r(t) + \Delta(\theta_{max} - \theta_{min})\cdot 10 \tag{9}$$

for position (x,y), velocity (u,v), and additional "noise" increments σ_*. The increments σ_* are a simple way to introduce additional time evolution and are generated independently and stochastically, with the random number generator seeded with 0 for reproducibility. During evolution, we bound $r \geq 10$ km and $\theta_{\max} \geq \theta_{\min} + 1$ K. By construction, a given disk should always and only correspond to itself over time.

For the control case, we initialize 50 objects with 20 km uniform x-y spacing. The initial velocities are sampled from uniform distributions $u_0, v_0 \in [-5, 5]$ m s^{-1}, core potential temperatures from a normal distribution $\theta \in N(275, 15)$ K, intensities from $\theta_{\max} - \theta_{\min} \in |N(0, 15)|$ K, and radii $r \in N(800, 100)$ km, motivated by Fig. 6 in Hakim and Canavan (2005). The noise terms sample centered uniform distributions of $\sigma_u, \sigma_v \in [-5, 5]$ m s^{-1} and $\sigma_\theta \in [-5, 5]$ K. Evolving each cell independently for twelve 6 h timesteps, we can consider the correspondences determined by each approach (Fig. 4). H05 and K10 create numerous false correspondences between nearby objects but miss no correspondences. The stricter 1-1 definition for major correspondences in TPVTrack greatly reduces the rate of false correspondences but yields several false negatives. Especially since the parameters for the control setup are somewhat arbitrary, several perturbations from the control have been tested. Changes to the initial separation, timestep, velocity amplitudes, and noise amplitudes (not shown) also yield results consistent with the following.

The simple test case illustrates characteristics of each approach. H05 creates correspondences whenever candidates fall within a distance cutoff. K10 creates correspondences when sufficient horizontal overlap occurs under advection. Correspondences from both are less robust for multiple close objects than TPVTrack, which incorporates horizontal overlap, vertical overlap, and 1-1 major correspondences. Rather than subjectively value one approach as absolutely better than another, these could be considered complementary tools. Overlap is useful for identifying possibly associated features through time. Further major correspondences use additional information to provide more targeting for following a distinct, individual feature.

3.2 Historical test cases

The purposefully simplified segmentation and correspondence test cases permit precise reference solutions. Further evaluation using more realistic historical data centers on two questions: are TPVs tracked by TPVTrack? Are tracks TPVs? Tracks were constructed using the default settings with $0.5^\circ \times 0.5^\circ$ ERA-Interim 2 PVU dynamic tropopause data for 1979 to 2015 every six hours. TPVs are defined as tracks with a core at genesis north of 60° N lasting at least 2 days. This relaxes the "Arctic" category of Hakim and Canavan (2005) to tracks that start polewards of the mean position of the polar jet.

3.2.1 Are long-lived TPVs tracked?

A track in 2006 is ~~the second~~ one of the longest in this dataset, from 06 July 2006 to 30 September 2006. It is of particular interest because of its long lifetime and associations with Arctic sea ice loss. Tracking manually, we can follow this TPV forming from filamentation and splitting on 03 July over western Siberia to dissipating on 30 September over eastern Siberia after traversing much of the central and eastern Arctic. The advective nature of the TPV is evident in comparing the TPV's track to a trajectory model (Stein et al., 2015). For example, 5 day forward and backward trajectories from 15 August 2006 00 UTC about the core using Global Data Assimilation System meteorology and model vertical velocity are near the core for the 10 day period. Using several less robust segmentation and correspondence settings, the TPV's track stops on 29 July, 03 September, 09 September, or 17 September instead. Figure 5 depicts the states driving these sensitivities. Generally, given multiple TPVs with overlap, major correspondences are discriminated quantitatively by similarity scores.

Ullrich and Zarzycki (2017) argue that a closed contour criterion is appealing since it is less sensitive to grid resolution relative to using only nearest neighbors. Here, we quantify some of TPVTrack's spatial sensitivities by segmenting an existing ensemble of full-physics, limited area simulations with model grid spacing from 12 to 120 km. We use WRF-ARW version 3.8.1 (WRF Users, 2016) on a roughly 6000 km by 6000 km domain centered on the north pole with 41 vertical levels up to 10 hPa with a time step of $\Delta t = 5\Delta x$ (e.g., 600 s for 120 km). Initial conditions on 15 August 2006 00 UTC are from GFS analysis with lateral boundary conditions updated every 3 h from alternating GFS 6 h analyses and 3 h forecasts. Physics parameterizations include Morrison 2-moment microphysics (Morrison et al., 2009), Rapid Radiative Transfer Model longwave radiation (RRTM-LW; Mlawer et al. (1997)), RRTM for General circulation models shortwave radiation (RRTMG-SW; Iacono et al. (2008)), Unified Noah land surface model (Tewari et al., 2004), and Yonsei University planetary boundary layer (Hong et al., 2006). The tropopause is diagnosed using NCL functions wrf_pvo to calculate potential vorticity and wrf_user_intrp3 to interpolate to 2 PVU, similar to values found by ~~trapping~~ interpolating in the vertical to 2 PVU by searching down from the model top (Sect. 3.2.3) for these grid scales~~(not shown)~~.

While the number of local extrema increases for the higher-resolution grids within the first day as the model spins up from the external initial conditions, the number of regional extrema remains nearly constant. Comparing the segmentations for the 12, 45, and 90 km grids after 3 days, there are combined effects of dynamical model and segmentation filters (Fig. 6). ~~The higher resolution grid has~~ Simulations using higher resolution grids have finer scale structures (Fig. 6.a,b,c) that are preserved in the ~~segmented objects~~ respectively segmented basins (Fig. 6.e,h,i). A larger filter radius for regional extrema reduces the number of basins and generally increases ~~their intensities~~ amplitudes (Fig. 6.e,g). Using the last closed contour over-restricts most of the basins (Fig. 6.d,e,f). Note that over-restriction can be basin-dependent, with the low near (1000, 3200) similar across all panels. Focusing on the long-lived 2006 TPV near (3000, 3000), the segmented basin is larger for larger cutoff radius, boundary contour percentile, or grid spacing (Fig. 7). The r1000 0% basin at 12 km is spuriously small. The large imposed filter generates few, large regional watershed basins. The minimum potential temperature along the resulting extended boundary is near that of the core, greatly reducing the restricted extent. Vortex cores may also be near the basin's boundary dynamically, like in conditions of strong deformation. In general, the shape of the distribution of potential temperature on the boundary of the watershed basin determines the sensitivity to the restriction percentile threshold; a uniform distribution with small range would exhibit little sensitivity.

The longest anticyclone track starts on 28 April 1979 and lasts 25 days ~~.~~ (Fig. 8). Synoptically, a filament from an amplifying Rossby wave break over the northeastern Pacific Ocean develops into a closed feature, couples with numerous cyclones over its entry into and traversal of the Arctic, and dissipates into the polar jet stream east of Greenland. TPVTrack's track exactly matches our manual track.

3.2.2 Bulk evaluation of tracks

Case studies are central to the iterative development of TPVTrack. A common sensitivity, as in the summer 2006 example, occurs when TPVs highly deform or intensify while in a neighborhood with multiple candidate overlaps. Beyond case studies and estimating skill through verification against a limited set of subjective and questionably reproducible manual tracks, bulk

statistics can be used to evaluate methods (Lakshmanan and Smith, 2010). What are the characteristics of tracked TPVs? Are these consistent with expectations? Note that a TPV climatology is beyond the scope of this paper and the focus of a separate work.

While a fuller understanding of the life cycles of TPVs is a matter of ongoing research, Hakim and Canavan (2005) shows TPVs to be generally smaller in radius at genesis or lysis in comparison to the rest of their lifetimes. Both cyclonic and anticyclonic tracked TPVs reach their minimum radius at the beginning or ends of tracks in the majority of cases (Fig. 9). The tail-dense distributions are also similar for time of minimum amplitude and circulation (not shown). Maximum radius tends to occur between genesis and lysis (not shown). Tracks with larger maximum radius have larger maximum circulation, amplitude, and lifetime ($p < 0.01$).

By construction, tracked TPVs are regional lows in potential temperature with cyclonic vorticity or highs in potential temperature with anticyclonic vorticity. To evaluate TPVs spatially, the Okubo-Weiss parameter (Okubo, 1970; Weiss, 1991) W is:

$$W = s_n^2 + s_s^2 - \zeta^2 \tag{10}$$

where $s_n = \frac{\partial u}{\partial x} - \frac{\partial v}{\partial y}$, $s_s = \frac{\partial v}{\partial x} + \frac{\partial u}{\partial y}$, and $\zeta = \frac{\partial v}{\partial x} - \frac{\partial u}{\partial y}$ are normal strain, shear strain, and relative vorticity, respectively. We include additional metric terms for latitude/longitude grids treating convergence near the poles (Saucier, 1955). If relative vorticity dominates strain, $W < 0$. Maps of W tend to exhibit three regions: rotational vortex cores characterized by negative W, strain regions surrounding vortices characterized by positive W, and background with small absolute W (Provenzale, 1999). For a TPV-relative reference frame, we center a cardinally-oriented computational grid with uniform 30 km spacing about each TPV core. Values within tracked basins are nearest-neighbor interpolated from the original grid onto the reference grid through a stereographic projection centered on the TPV core. Given the expense of remapping following each TPV, we consider one year (Fig. 10.a,b). Cyclones are frequently rotational about the core, while for anticyclones the region of concentrated rotation is less frequent, broader, and offset northwards of the core. Considering tracks with at least a 10 days lifetime yields a more rotationally dominated structure about the core (not shown). Geographically, (Fig. 10.c,d), tracked cyclones tend to have negative W except near orographic features, while tracked anticyclones appear to have higher strain at lower latitudes where shear interactions with jet streams would occur.

Thus, in bulk, TPVs from TPVTrack are consistent with eddies with rotational cores exhibiting reasonable genesis and lysis and consistency between metrics. The symmetric, core-relative vortex structure is less frequent for features that are shorter-lived, over mountainous terrain for cyclones, or interacting with lower latitudes for anticyclones. However, it is unclear how discriminating these bulk results are. An intercomparison of trackers would further inform relative algorithm evaluation.

Most likely occurrences of TPVs from TPVTrack in the central Arctic and Canadian Archipelago (Fig. 11.a,b) are consistent with Cavallo and Hakim (2010). Varying the geographic criteria for defining TPVs from tracks yields differences of the same order as the expected density for TPVs near mid-latitudes and little difference polewards (Fig. 11.c). Decreasing the minimum duration required of tracks (Fig. 11.d) has a relatively smaller impact, generally increasing the total number of TPVs with maximum differences broadly over the North Atlantic and European Arctic.

3.2.3 Accommodations for model data

Several practical issues have been noted through application of TPVTrack to model data, largely related to the diagnosis of the tropopause and scales of interest. Treatment of missing data via iterative flood-fill (Sect. 2.2.1) is included in the code release. Especially at finer grid scales, relative vorticity may need to be filtered, similar to the "noise" issues seen in the idealized segmentation test case when assigning cells to either highs or lows. In computing metrics, raw or filtered fields could be used, in a trade-off of consistency with the full flow field versus scales of interest.

While a dataset may define its own diagnosis of the dynamic tropopause, conventionally, the dynamic tropopause is found by searching down from some upper level or model top for a potential vorticity threshold within each column independently, possibly with thickness criteria (e.g., Zängl and Hoinka (2001)). However, vertical mixing, folding, and convective-scale vorticity can generate ambiguity with multiple candidates for the tropopause within a column, occurring globally (Añel et al., 2008). Beyond thickness conditions, we have found the tropopause as a surface to be identified more robustly by including consistency with information outside the column. To identify the tropopause as a layer between the respectively connected high potential vorticity stratosphere and low potential vorticity troposphere, flood-fill segmentations of the stratosphere and troposphere can be performed. For example, starting with initial seed cells near the surface with low potential vorticity, all connected values below a chosen potential vorticity threshold can be found by iterating through valid nearest neighbors. This identification of the connected troposphere yields the tropopause as the top of the tropospheric volume. Note that the top of the troposphere is above the bottom of the stratosphere, and the bottom of the stratosphere may reach the surface, especially within potential vorticity towers ~~.~~ (e.g., Pang and Fu (2017), Fig. 6c). If model data is archived with sparse vertical levels, this diagnosis may have larger uncertainties. Also note that the chosen ~~PVU~~ isertelic level can impact TPV metrics quantitatively (e.g., Fig. 4 in Cavallo and Hakim (2009)).

Typical model output frequencies range from hourly to daily. The simple advection, constant amplitude overlap similarity is less appropriate with decreasing time resolution. Tracking on under-resolved fields may increase systematic misidentifications of TPVs. Note that genesis at the initial time and lysis at the final time are not dependent on the flow field but rather on truncated time input.

4 Conclusions

TPVTrack is a software package for ~~automated~~ rule-based tracking of TPVs motivated ~~by~~ to address limitations of spatial characterization and time tracking within existing approaches and to further our understanding of TPVs and their connections throughout the Earth system. The two primary goals are more robust representations of (1) the spatial structure of features, through restricted regional watershed basins, and (2) mergers and splits in time, through major correspondences from 1-1 similarity overlap. From the simplified test cases, robust segmentation is directly dependent on the vertical vorticity used to associate cells to cyclones and anticyclones, possibly requiring smoothing for noisy fields in pre-processing. Major correspondences reduce the false positive associations but can increase the false negative rate, relative to H05 and K10. Horizontal overlap may benefit from more frequent output or a more sophisticated trajectory model. Including information from more

times may make tracking more robust, say with further hypothesis management (e.g., Reid (1979)). Historical test cases reveal larger uncertainties in tracks and sensitivities to user-defined parameters when multiple TPVs are in a neighborhood.

~~There has yet to be a methodological study considering whether the approach used to define TPVs impacts~~ Using the default settings, TPVs are most often found over the central Arctic and Canadian Archipelago, decreasing in density with distance. In bulk, cyclone structure is frequently characterized by a rotationally dominant core, while anticyclones are less so. Cores are less rotationally dominant for shorter lifetimes, near mountains for cyclones, or near lower latitudes for anticyclones. A TPV's smallest radius tends to occur at the beginning or end of its life, while maximum occurs between. Individual TPVs can live several months, but lifetime, size, and climatological density can exhibit sensitivities and corresponding uncertainty to input data, tracking method, and TPVTrack settings. TPVTrack has four user-specified parameters. Size of the neighborhood for regional extrema and percentile for restriction of watershed basins are specific to TPVTrack. Minimum lifetime and polar criteria for tracks are applicable to all methods. Given these sensitivities, there are concerns over use of a single approach depending on the intended application.

Further analysis is important to distinguish the TPV-related information as opposed to automated, rule-based objects and whether our physical understanding is impacted. With complementary approaches, several algorithms may be combined constructively to reduce artifacts of any single approach during analysis. In addition to providing another method, TPVTrack's modular design and exposed settings file facilitate intra-algorithm sensitivity studies~~, e. g., varying the restriction percentile or radius for testing regional extrema. Also, an alternate method~~. Note that alternate methods for segmentation (e.g., connected anomalies of minimum intensity or size) could be ~~used as a different model~~ incorporated within TPVTrack as different models for spatial shape, with no modifications needed for the rest of the algorithm. Perhaps more fundamentally, can a TPV have multiple centers (e.g., Hanley and Caballero (2012))? Users focusing on case studies are encouraged to test robustness of individual tracks, while systematic artifacts may be more relevant for climatologies, including sensitivity to the input data.

The equivalent radius of a TPV's basin measures one aspect of size. Physically, the size of a TPV may refer to a number of scales, including where fluid is more trapped near the core, the (anti)cyclonic basin about an extremum, minimum distance of the core to a boundary, or the region of flow influenced by the potential vorticity anomaly. Are there clear relationships between these scales? Are different measures more meaningful or useful for different applications? TPVTrack provides one component towards addressing this question.

Towards better understanding of the vertical structure of TPVs, tracks can be constructed on surfaces other than the tropopause. Associating the separate objects between levels then stitches together the evolution of spatially 3-dimensional objects over time. For a more integrated approach, it may be possible to construct spatially 3-dimensional objects directly by defining appropriate scalings or anomalies for steepest descent in the watershed segmentation and contour restriction. Similar overlap correspondence could then be used for tracking.

Towards tracking features other than TPVs, broadly, TPVTrack identifies coherent objects and creates tracks by overlap similarity over time, in effect defining the spatial structure and time evolution. Approaches are under development for jet streaks, surface cyclones, cold air outbreaks, and sea ice loss. To allow more flexibility in input data, it may be possible to track on a masked surface where not all values are valid. An application would be to alleviate the uncertainties in estimating

sea level pressure over high terrain (e.g., Pauley (1998)) for surface cyclone tracking. Evaluation is needed to inform how the tracked objects map to the physical features for each application, to maximize interpretation of physical results rather than output from ~~automated~~ rule-based algorithms. A general and comprehensive method for evaluation is elusive. Given tracks of multiple features, a central problem is to define and justify when they interact.

Code availability. The software can be obtained publicly at https://github.com/nickszap/tpvTrack (doi:10.5281/zenodo.1311001) or from the authors by e-mail.

Data availability. The datasets and software are publicly available. ERA-Interim variables can be obtained from the ECMWF data server (https://software.ecmwf.int/wiki/display/WEBAPI/Access+ECMWF+Public+Datasets). The WRF model is available from the WRF Users' page (http://www2.mmm.ucar.edu/wrf/users/), and GFS initial and boundary conditions are available from the National Centers for Environmental Information (https://www.ncdc.noaa.gov/data-access/model-data/model-datasets/global-forcast-system-gfs). In addition, simulation output is available from the corresponding author on request.

Author contributions. Both authors have contributed to tracker development, evaluation, and writing of the paper.

Competing interests. The authors declare that they have no conflicts of interest.

Acknowledgements. The authors thank the members of the University of Oklahoma Arctic and Antarctic Research Group for discussion, especially Christopher Riedel for running the WRF simulations. The authors thank topical editor Juan Antonio Añel and reviews from Baird Langenbrunner and an anonymous referee for providing constructive comments. This work has been supported by the Office of Naval Research award N00015-1-2220.

References

Añel, J. A., Antuña, J. C., de la Torre, L., Castanheira, J. M., and Gimeno, L.: Climatological features of global multiple tropopause events, J. Geophys. Res., 113, 2008.

Béguin, A., Martius, O., Sprenger, M., Spichtinger, P., Folini, D., and Wernli, H.: Tropopause level Rossby wave breaking in the Northern Hemisphere: a feature-based validation of the ECHAM5-HAM climate model, Int. J. Climatol., 2012.

Bluestein, H. B.: Synoptic-dynamic Meteorology in Midlatitudes: Observations and theory of weather systems, vol. 2, Taylor & Francis, 1992.

Cavallo, S. M. and Hakim, G. J.: Potential vorticity diagnosis of a tropopause polar cyclone, Mon. Weather Rev., 137, 1358–1371, 2009.

Cavallo, S. M. and Hakim, G. J.: Composite structure of tropopause polar cyclones, Mon. Weather Rev., 138, 3840–3857, 2010.

Dalcín, L., Paz, R., Storti, M., and D'Elía, J.: MPI for Python: Performance improvements and MPI-2 extensions, Journal of Parallel and Distributed Computing, 68, 655–662, 2008.

Davis, C. A. and Emanuel, K. A.: Potential vorticity diagnostics of cyclogenesis, Mon. Weather Rev., 119, 1929–1953, 1991.

Dee, D., Uppala, S., Simmons, A., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M., Balsamo, G., Bauer, P., et al.: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system, Quart. J. Roy. Meteorol. Soc., 137, 553–597, 2011.

Gettelman, A., Hoor, P., Pan, L., Randel, W., Hegglin, M. I., and Birner, T.: The extratropical upper troposphere and lower stratosphere, Rev. Geophys., 49, 2011.

Grams, C. M., Wernli, H., Böttcher, M., Čampa, J., Corsmeier, U., Jones, S. C., Keller, J. H., Lenz, C.-J., and Wiegand, L.: The key role of diabatic processes in modifying the upper-tropospheric wave guide: a North Atlantic case-study, Quart. J. Roy. Meteorol. Soc., 137, 2174–2193, 2011.

Hakim, G. J.: Climatology of coherent structures on the extratropical tropopause, Mon. Weather Rev., 128, 385–406, 2000.

Hakim, G. J. and Canavan, A. K.: Observed cyclone–anticyclone tropopause vortex asymmetries., J. Atmos. Sci., 62, 2005.

Hanley, J. and Caballero, R.: Objective identification and tracking of multicentre cyclones in the ERA-Interim reanalysis dataset, Quart. J. Roy. Meteorol. Soc., 138, 612–625, 2012.

Hewson, T. D. and Titley, H. A.: Objective identification, typing and tracking of the complete life-cycles of cyclonic features at high spatial resolution, Meteorological Applications, 17, 355–381, 2010.

Highwood, E. and Hoskins, B.: The tropical tropopause, Quart. J. Roy. Meteorol. Soc., 124, 1579–1604, 1998.

Hodges, K.: Adaptive constraints for feature tracking, Mon. Weather Rev., 127, 1999.

Holton, J. R. and Hakim, G. J.: An introduction to dynamic meteorology, Academic press, 2013.

Hong, S.-Y., Noh, Y., and Dudhia, J.: A new vertical diffusion package with an explicit treatment of entrainment processes, Mon. Weather Rev., 134, 2318–2341, 2006.

Horn, M., Walsh, K., Zhao, M., Camargo, S. J., Scoccimarro, E., Murakami, H., Wang, H., Ballinger, A., Kumar, A., Shaevitz, D. A., et al.: Tracking scheme dependence of simulated tropical cyclone response to idealized climate simulations, J. Climate, 27, 9197–9213, 2014.

Hoskins, B. J., McIntyre, M., and Robertson, A. W.: On the use and significance of isentropic potential vorticity maps, Quart. J. Roy. Meteorol. Soc., 111, 877–946, 1985.

Iacono, M. J., Delamere, J. S., Mlawer, E. J., Shephard, M. W., Clough, S. A., and Collins, W. D.: Radiative forcing by long-lived greenhouse gases: Calculations with the AER radiative transfer models, J. Geophys. Res., 113, 2008.

Ivanova, A.: The tropopause: Variety of definitions and modern approaches to identification, Russian Meteorology and Hydrology, 38, 808, 2013.

Kew, S. F., Sprenger, M., and Davies, H. C.: Potential Vorticity Anomalies of the Lowermost Stratosphere: A 10-Yr Winter Climatology, Mon. Weather Rev., 138, 2010.

Lakshmanan, V. and Smith, T.: An objective method of evaluating and devising storm-tracking algorithms, Weather Forecast., 25, 701–709, 2010.

Limbach, S., Schömer, E., and Wernli, H.: Detection, tracking and event localization of jet stream features in 4-D atmospheric data, Geosci. Model Dev., 5, 457–470, 2012.

Masarik, M. T. and Schubert, W. H.: Analytical solutions of the potential vorticity invertibility principle, Journal of Advances in Modeling Earth Systems, 5, 366–381, 2013.

Mlawer, E. J., Taubman, S. J., Brown, P. D., Iacono, M. J., and Clough, S. A.: Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave, J. Geophys. Res., 102, 16 663–16 682, 1997.

Morgan, M. C. and Nielsen-Gammon, J. W.: Using tropopause maps to diagnose midlatitude weather systems, Mon. Weather Rev., 126, 2555–2579, 1998.

Morrison, H., Thompson, G., and Tatarskii, V.: Impact of cloud microphysics on the development of trailing stratiform precipitation in a simulated squall line: Comparison of one-and two-moment schemes, Mon. Weather Rev., 137, 991–1007, 2009.

Neu, U., Akperov, M. G., Bellenbaum, N., Benestad, R., Blender, R., Caballero, R., Cocozza, A., Dacre, H. F., Feng, Y., Fraedrich, K., et al.: IMILAST: A community effort to intercompare extratropical cyclone detection and tracking algorithms, Bull. Amer. Meteorol. Soc., 94, 529–547, 2013.

Nieto, R., Sprenger, M., Wernli, H., Trigo, R., and Gimeno, L.: Identification and climatology of cut-off lows near the tropopause, Annals of the New York Academy of Sciences, 1146, 256–290, 2008.

Okubo, A.: Horizontal dispersion of floatable particles in the vicinity of velocity singularities such as convergences, Deep sea research and oceanographic abstracts, 17, 445–454, 1970.

Pang, H. and Fu, G.: Case study of potential vorticity tower in three explosive cyclones over Eastern Asia, J. Atmos. Sci., 74, 1445–1454, 2017.

Pauley, P. M.: An example of uncertainty in sea level pressure reduction, Weather and forecasting, 13, 833–850, 1998.

Pedlosky, J.: Geophysical fluid dynamics, Springer Science & Business Media, 2013.

Provenzale, A.: Transport by coherent barotropic vortices, Annual review of fluid mechanics, 31, 55–93, 1999.

Reid, D.: An algorithm for tracking multiple targets, IEEE transactions on Automatic Control, 24, 843–854, 1979.

Rew, R. and Davis, G.: NetCDF: an interface for scientific data access, IEEE computer graphics and applications, 10, 76–82, 1990.

Roache, P. J.: Verification and validation in computational science and engineering, Hermosa Albuquerque, 1998.

Roerdink, J. B. and Meijster, A.: The watershed transform: Definitions, algorithms and parallelization strategies, Fundamenta Informaticae, 41, 187–228, 2000.

Rossby, C.-G.: Relation between variations in the intensity of the zonal circulation of the atmosphere and the displacements of the semi-permanent centers of action, Journal of Marine Research, 2, 38–55, 1939.

Saha, S., Moorthi, S., Pan, H.-L., Wu, X., Wang, J., Nadiga, S., Tripp, P., Kistler, R., Woollen, J., Behringer, D., et al.: The NCEP climate forecast system reanalysis, Bull. Amer. Meteorol. Soc., 91, 1015–1057, 2010.

Saucier, W. J.: Principles of meteorological analysis, University of Chicago Press, 1955.

Serra, J.: Image analysis and mathematical morphology, Academic Press, Inc., 1983.

Shapiro, M., Hampel, T., and Krueger, A.: The Arctic tropopause fold, Mon. Weather Rev., 115, 444–454, 1987.

Shields, C. A., Rutz, J. J., Leung, L.-Y., Ralph, F. M., Wehner, M., Kawzenuk, B., Lora, J. M., McClenny, E., Osborne, T., Payne, A. E., et al.: Atmospheric River Tracking Method Intercomparison Project (ARTMIP): project goals and experimental design, Geosci. Model Dev., 11, 2455–2474, 2018.

Simmonds, I. and Rudeva, I.: The great Arctic cyclone of August 2012, Geophys. Res. Lett., 39, 2012.

Simmonds, I. and Rudeva, I.: A comparison of tracking methods for extreme cyclones in the Arctic basin, Tellus A, 66, 2014.

Skamarock, W. C., Klemp, J. B., Dudhia, J., Gill, D. O., Barker, D. M., Wang, W., and Powers, J. G.: A description of the advanced research WRF version 2, Tech. rep., NCAR, 2005.

Skamarock, W. C., Klemp, J. B., Duda, M. G., Fowler, L. D., Park, S.-H., and Ringler, T. D.: A multiscale nonhydrostatic atmospheric model using centroidal Voronoi tesselations and C-grid staggering, Mon. Weather Rev., 140, 3090–3105, 2012.

Sprenger, M., Wernli, H., and Bourqui, M.: Stratosphere-troposphere exchange and its relation to potential vorticity streamers and cutoffs near the extratropical tropopause, J. Atmos. Sci., 64, 1587–1602, 2007.

Stein, A., Draxler, R. R., Rolph, G. D., Stunder, B. J., Cohen, M., and Ngan, F.: NOAA's HYSPLIT atmospheric transport and dispersion modeling system, Bull. Amer. Meteorol. Soc., 96, 2059–2077, 2015.

Tebaldi, C. and Knutti, R.: The use of the multi-model ensemble in probabilistic climate projections, Philos. Trans. Roy. Soc. London, 365, 2053–2075, 2007.

Tewari, M., Chen, F., Wang, W., Dudhia, J., LeMone, M., Mitchell, K., Ek, M., Gayno, G., Wegiel, J., and Cuenca, R.: Implementation and verification of the unified NOAH land surface model in the WRF model, in: 20th conference on weather analysis and forecasting/16th conference on numerical weather prediction, vol. 1115, 2004.

Thorpe, A. J.: Synoptic scale disturbances with circular symmetry, Mon. Weather Rev., 114, 1384–1389, 1986.

Tilinina, N., Gulev, S. K., Rudeva, I., and Koltermann, P.: Comparing cyclone life cycle characteristics and their interannual variability in different reanalyses, J. Climate, 26, 6419–6438, 2013.

Ullrich, P. A. and Zarzycki, C. M.: TempestExtremes v1.0: A framework for scale-insensitive pointwise feature tracking on unstructured grids, Geosci. Model Dev., 10, 1069–1090, 2017.

Walsh, K., Fiorino, M., Landsea, C., and McInnes, K.: Objectively determined resolution-dependent threshold criteria for the detection of tropical cyclones in climate models and reanalyses, J. Climate, 20, 2307–2314, 2007.

Weiss, J.: The dynamics of enstrophy transfer in two-dimensional hydrodynamics, Physica D: Nonlinear Phenomena, 48, 273–294, 1991.

Whitaker, J.: netCDF4 API documentation, http://unidata.github.io/netcdf4-python/, 2015.

WRF Users: WRF Model Version 3.8.1: UPDATES, http://www2.mmm.ucar.edu/wrf/users/wrfv3.8/updates-3.8.1.html, 2016.

Yamazaki, A., Inoue, J., Dethloff, K., Maturilli, M., and König-Langlo, G.: Impact of radiosonde observations on forecasting summertime Arctic cyclone formation, J. Geophys. Res., 120, 3249–3273, 2015.

Zängl, G. and Hoinka, K. P.: The tropopause in the polar regions, J. Climate, 14, 3117–3139, 2001.

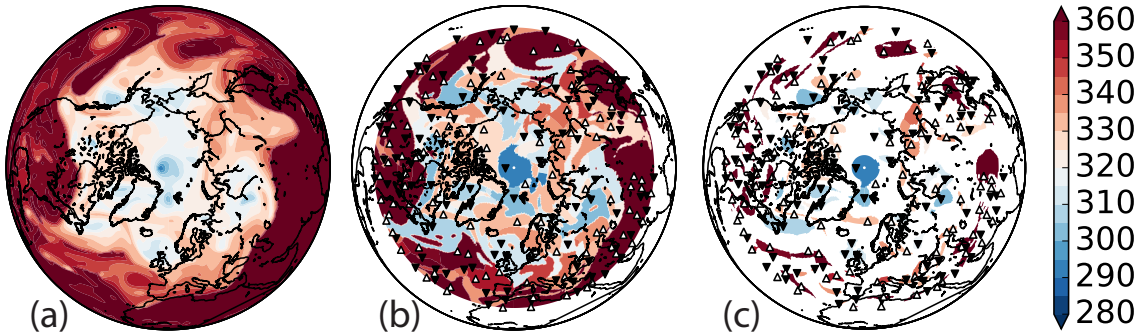

Figure 1. (a) Tropopause potential temperature (K; 5 K contour interval) from ERA-Interim on 01 August 2006 00 UTC. (b) Segmentation into low (cyclonic) and high (anticyclonic) regional watershed basins north of 30° N. Regional extrema for cyclones and anticyclones are depicted by black downward and white upward triangle markers, respectively. (c) as in (b), with restriction within basins to the 10^{th} percentile of boundary amplitudes.

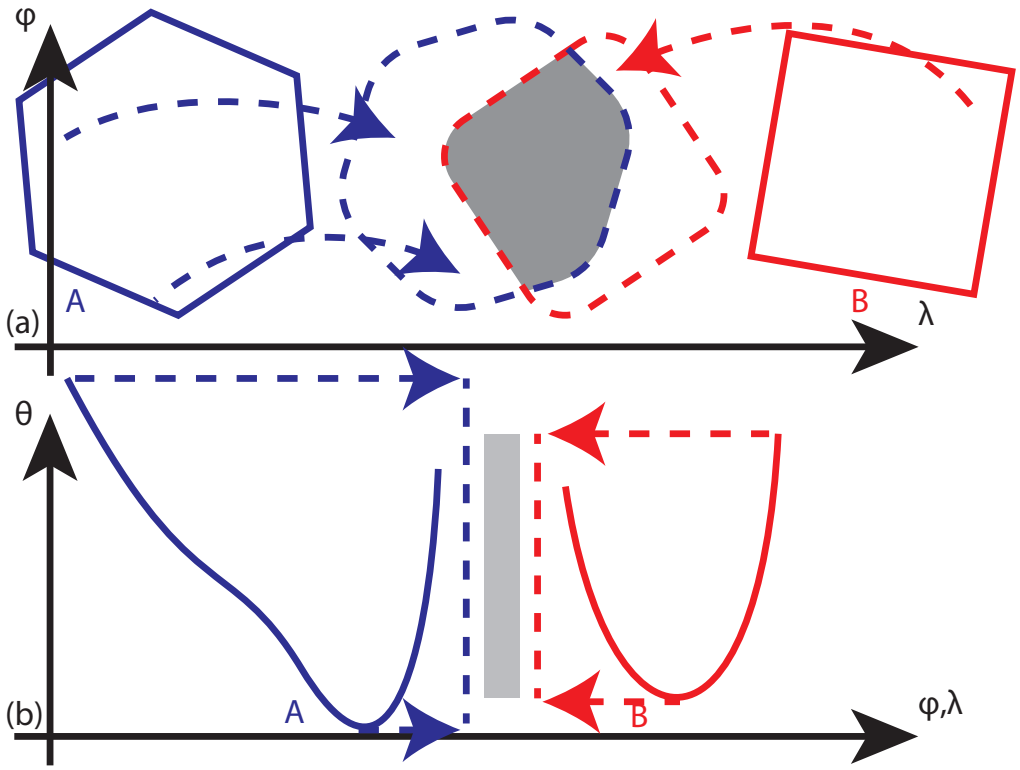

Figure 2. Schematic for the ~~two~~ (a) horizontal and (b) vertical components of time correspondence similarity quantified by overlap (gray) of ~~(~~a ~~) horizontal cell-to-cell advection and (b) persistent vertical potential temperature ranges for a~~ basin A at t_0 (blue) and basin B at $t_0 + \Delta t$ (red) at the intermediate half-time (dashed). (a) For the horizontal, each cell in the basin is advected, and overlap occurs in the common cells. (b) For the vertical, the potential temperature range $[\theta_{min}, \theta_{max}]$ of each basin is held constant, and overlap is the one-dimensional intersection of the ranges.

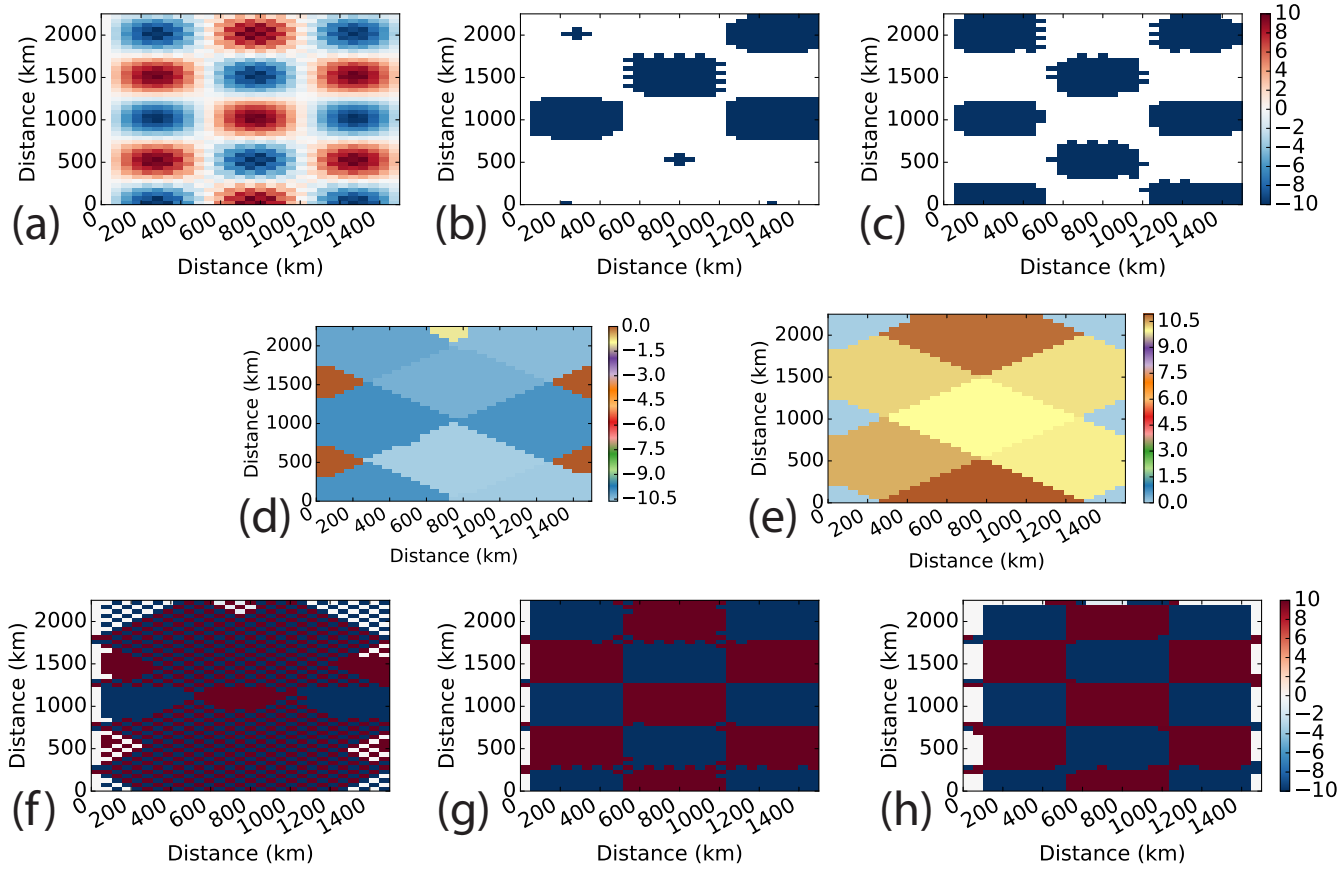

Figure 3. Comparison of (b) H05, (c) K10, and watershed (f,g,h) segmentations for an analytic surface with multiple scales (a; Equation 4). For the watershed method, each cell is in both a (d) low and (e) high regional watershed basin. Assignment to one is by the local relative vorticity, calculated by (f) centered finite differences on the raw field (g) a smoother Gaussian Laplacian with standard deviation of 1, and (h) sign of the vorticity is minus the sign of the surface (a). Each object in a segmentation (b-h) is colored by the intensity of its core. Colorbars (K) for (a,b,c,f,g,h) are the same.

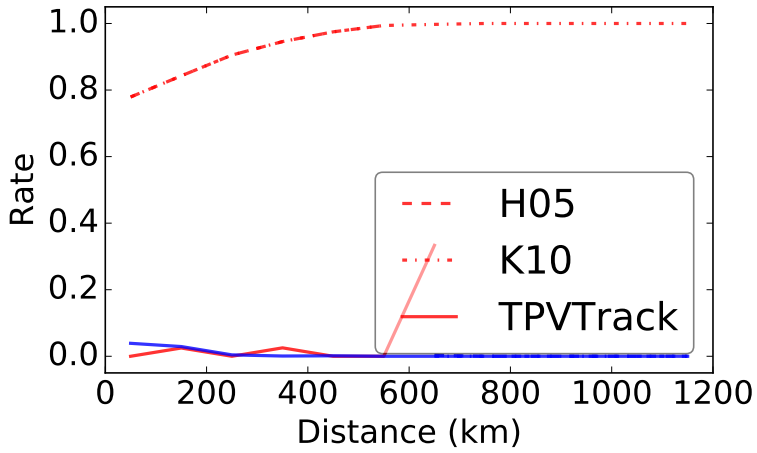

Figure 4. False positive (red) and negative (blue) time correspondence rates for control setup of idealized correspondence test case for H05 (dashed), K10 (dash-dotted), and TPVTrack (solid) approaches. Rates are in bins of 100 km core-core distance. For H05 and K10, false negatives are zero and false positives overlap for distances less than 600 km.

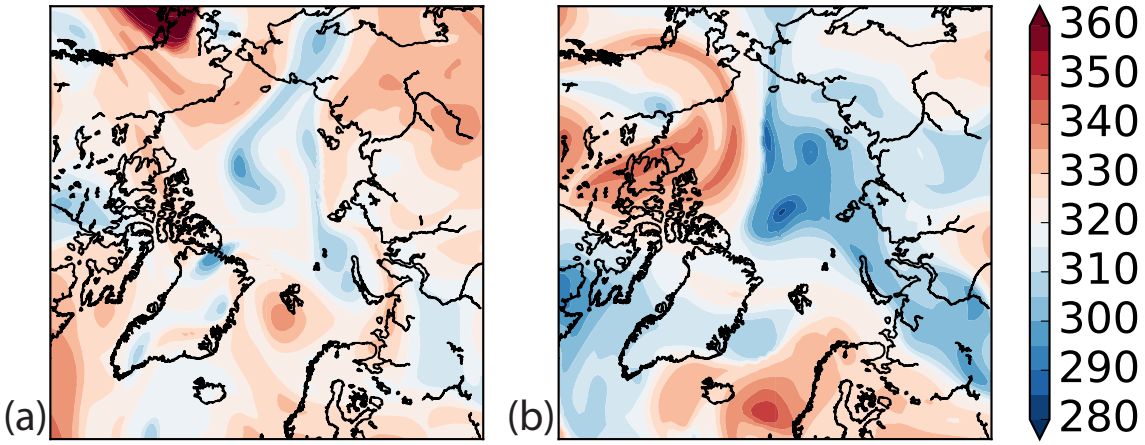

Figure 5. Maps of tropopause potential temperature (K; 5 K contour interval) for dates with sensitivities in long-lived Summer 2006 track, (a) 29 July 2006 00 UTC, (b) 17 September 2006 00 UTC

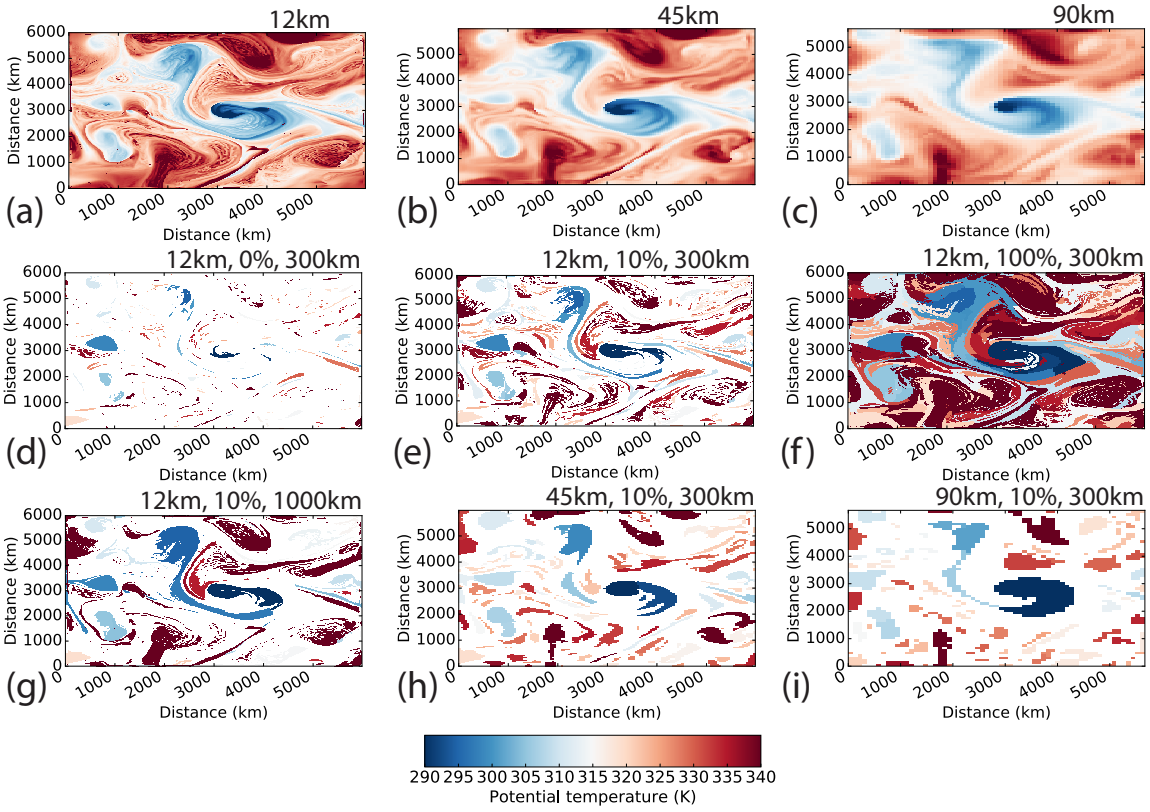

Figure 6. Segmentation of (a) 12 km, (b) 45 km, and (c) 90 km WRF tropopause (after 3 simulation days) for (d) 12 km, 0% restriction of boundary amplitudes, and 300 km regional extrema, (e) as in d, but 10% restriction, (f), as in d, but 100% restriction (g) as in e, but 1000 km regional extrema, (h) as in e, but for 45 km simulation, (i) as in e, but for 90 km simulation

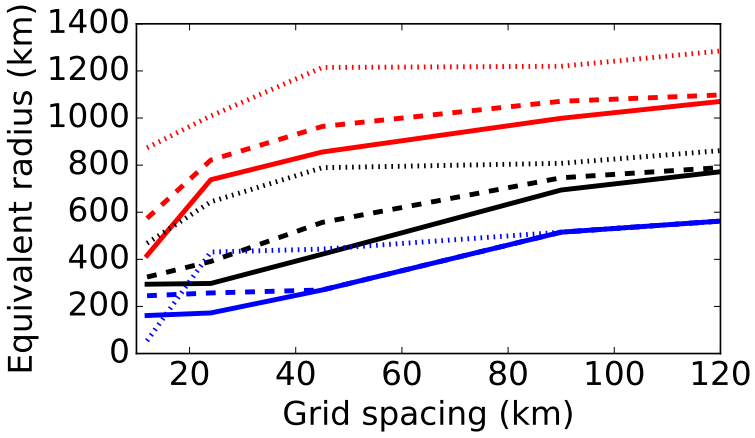

Figure 7. Equivalent radius of long-lived 2006 TPV from watershed segmentation of 12 to 120 km WRF simulations (at 3 days) with 300 (solid), 500 (dashed), and 1000 km (dotted) regional extrema and basins restricted to 0% (blue), 10% (black), and 100% (red) boundary amplitudes.

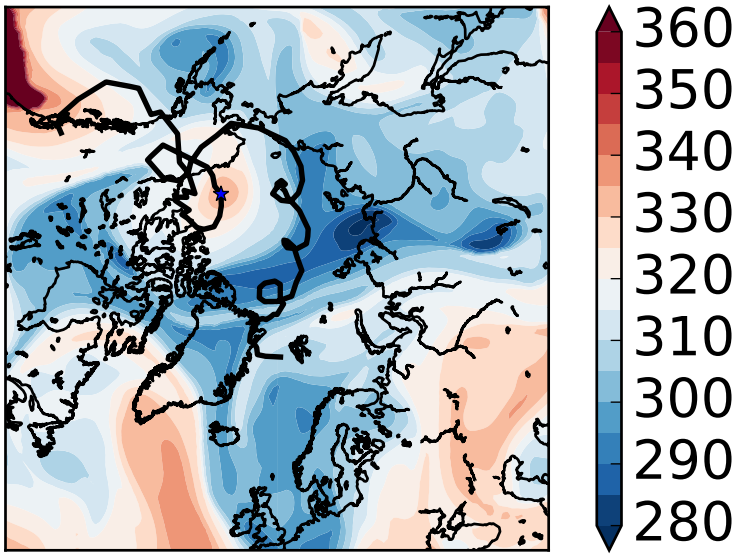

Figure 8. ~~Probability of~~ For the ~~normalized lifetime~~ longest anticyclone track in 1979 to 2015, map of dynamic tropopause potential temperature (~~0~~ K; 5 K interval) at ~~genesis~~ 03 May 1979 00 UTC, and ~~1 at lysis within each~~ track ~~)~~ of ~~minimum equivalent radius for cyclonic~~ the core overlaid in black. The location of the core (~~blue~~north of Alaska) ~~and anticyclonic (red) TPVs in 1979 to 2015.~~is starred at the common time.

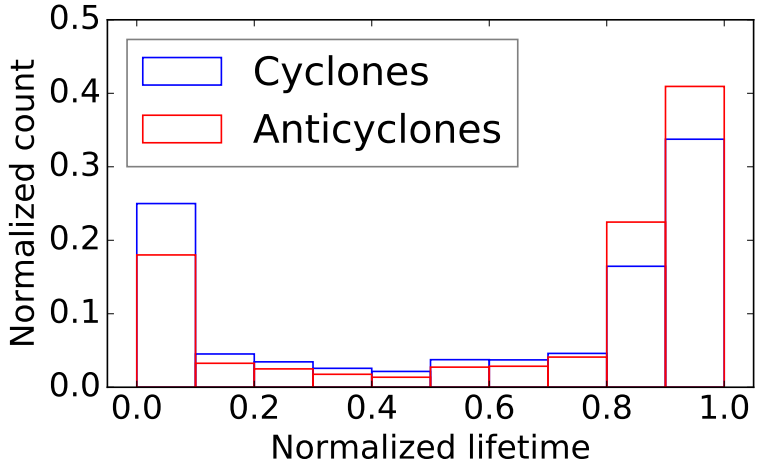

Figure 9. Fraction of all cyclonic (blue) and anticylonic (red) TPVs in 1979 to 2015 being their minimum equivalent radius as a function of stage between genesis and lysis. The normalized lifetime linear in time between 0 at genesis and 1 at lysis within each track. The count is normalized to a fraction by the total number of TPVs (41190 and 33392 for cyclones and anticyclones, respectively) and aggregated in 10 equally-spaced normalized lifetime bins.

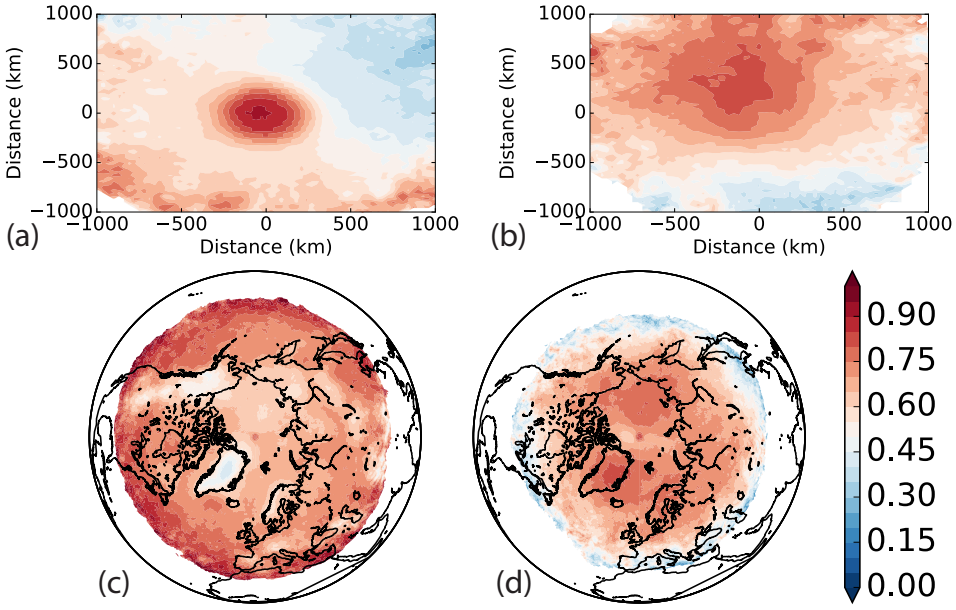

Figure 10. Fraction of time with negative Okubo-Weiss given a TPV present, (a) cyclonic TPVs in 2006 in a core-relative reference frame, (b) anticyclonic TPVs in 2006 in a core-relative reference frame, (c) cyclonic TPVs in 1979 to 2015 geographically, and (d) anticyclonic TPVs in 1979 to 2015 geographically. Values with fewer than 5 days of samples (on the domain boundaries) are masked in white. The contour interval is 5%.

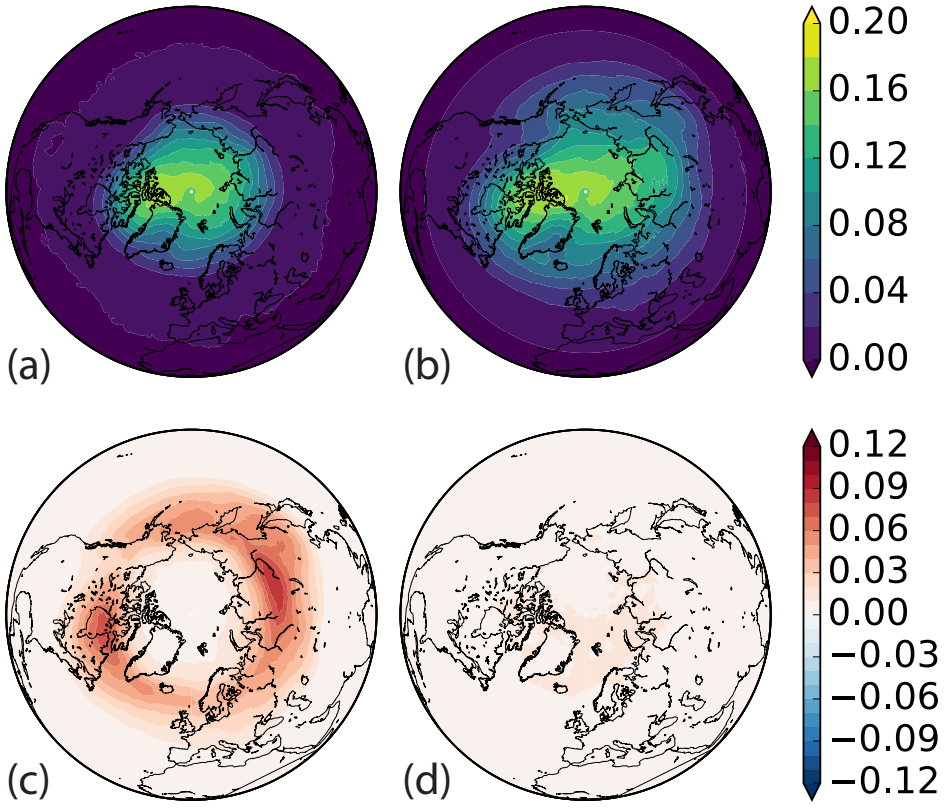

Figure 11. Geographic probability of a cyclonic TPV present from ERA-Interim 1979 to 2015, calculated by aggregating times within basins of tracks satisfying TPV criteria. (a) TPVs defined via H05 "Arctic" category, where TPVs are tracks that last at least 2 days with over 60% of their lifetimes north of 65° N. (b) Default TPVTrack settings, where TPVs are tracks lasting at least 2 days with genesis north of 60° N. (c) b-a. Difference in probability using different latitude criteria for defining TPVs of genesis north of 60° N minus 60% of lifetime north of 65° N. (d) Difference in probability using different lifetime criteria for defining TPVs of minimum lifetime of 1 day minus 2 days. A 300 km radius for regional extrema and 5% watershed boundary amplitude restriction apply to all. Contour intervals are 2% for a,b and 1% for c,d, respectively.