Response to Referee #2

July 9, 2020

Many thanks for your time and efforts on this manuscript. I'm giving a one-to-one response to your comments below.

Major comments

Line 153-154: Please clarify when the 800 km threshold is applied and when the 2000km is applied. How is the genesis stage of strong ARs defined?

All detection results are using a 2000 km length requirement except for the tracking across time section, and we mentioned at L365-366 that 800 km is used but it is required that the same AR has to reach at least 2000 km for at least once during its life time.

The sentences are changed to make this clearer:

> A typical 2000 km minimum length requirement is adopted to be consistent with previous studies, however, a relaxed 800 km threshold is also applied when we track ARs across time. This allows weaker systems, many of which occur during the genesis stage (the first time point of an AR track after tracked the same AR entity across time) of strong ARs, to be included, and helps depict a more complete picture of AR life cycle. However, it is required that the same AR reaches 2000 km or above at least one time step during its life time.

Line 153: Why the ARs with too large area or too large length need to be eliminated? Please provide more examples to demonstrate why the large scale ARs do not conform to the definition of AR. The existence of large scale ARs wound have different reasons, for example two ARs are merged mistakenly and become a large scale AR. It is not appropriated to eliminate them directly.

The maximum area/length requirements are set to fairly large values. The maximum length is set to 11,000 km, which is longer than the great circle distance (through Pacific, $\sim$ 10400 km) from Hong Kong ($\sim 22.3°N, 114.2°E$) to Seattle ($\sim 47.6°N, -122.3°W$). Assuming an average width of 700 km (which is on the larger side for typical ARs), this corresponds to an area of $\sim 7.3 \times 10^6 km^2$, which is smaller than the maximum area requirement set here ($10 \times 10^6 km^2$). These requirements are far larger than what could be expected from real world AR sizes. Therefore if they ever filtered out anything, it is pretty safe to say that those are not ARs.

Such cases only happen when two or more ARs are merged together, or some of the ARs are connected with the large and continuous moisture plumes in the tropics to form an area/length so large as to be filtered out. This happens more frequently when the structuring element E is set too large, as explained in Line 244-246. Figure 1 gives one such example. In Figure 1a, the structuring element is set to t7-s6. This merges the Pacific ARs with tropical plumes to form a big connected region that later gets filtered out. (b) is using the recommended t4-s6 parameters, this time the mid-latitude signals are better separated from tropical ones so the Pacific AR is correctly retained. Those small noisy contours will be filtered by the minimum area requirement.

Such gigantic regions should not be regarded as valid detections, as least by our standard, because if so, an easy and lazy solution would be setting an low enough IVT threshold, for instance $IVT \geq 50kg/m/s$, to guarantee the coverage of every AR possible at any instance. However, the results will not be super useful, the size is bound to be overestimated, the axis will follow weird paths. When used as inputs for subsequent analyses, they will not contribute positively to the analysis of the possible relationship with extra-tropical cyclones, the AR structures, the life cycle evolution, the duration of continent landfalling and precipitation impacts ... I would rather treat them as outliers and discard from the analyses.

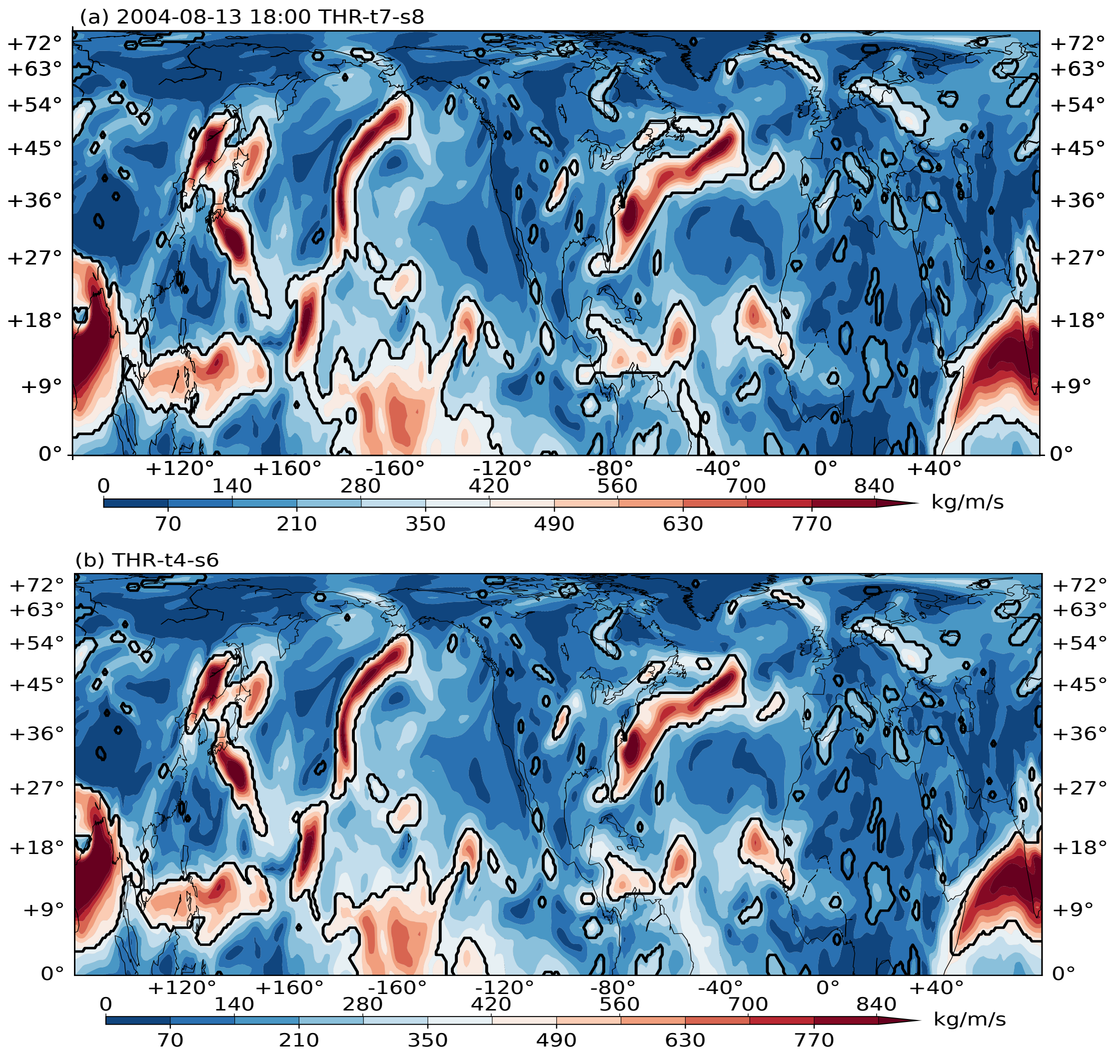

Figure 1: Northern Hemisphere IVT distribution at 2004-08-13 18 UTC. Black contours denote connected regions where the THR anomalies are greater than zero. (a) is the result using THR-t7-s8 parameters, (b) is the THR-t4-s6 parameters.

That being said, the current method does not handle the separation of merged ARs. Around a merging/splitting event, it can get very ambiguous when it should be treated as one AR or two. We did develop a method based on a metric in the field of topography to separate merged ARs. However, at the time of writing the method was not fully ready. The code (in progress) was included in the code we published, but the functionality was not turned on when producing the results. Therefore, only the very extreme cases that a single maximum length/area filtering can easily handle are treated as such.

In Figure 2 we give an example of the separation method. The ARs in question are in Southern Hemisphere, at $\sim 120°W$, 1980-01-06 09 UTC. Animation around this time point suggests that a side-to-side merging is happening. At this chosen time point, the 2 merging plumes leave a tiny gap in between them. When taking the non-zeros regions from the THR anomaly field as stated in the manuscript, they would be grouped into a single contour, but the separation method breaks it into two ARs. One time step later (3 hr using the MERRA2 reanalysis data), they are treated as a single AR and a merging happens. Again, when to decide a merging/splitting can be a subtle choice, but for many other situations when the grouped ARs are pretty far apart, the separation is quite unambiguous. We are planning to properly introduce this method in another paper, then the maximum length/area requirements are no longer needed.

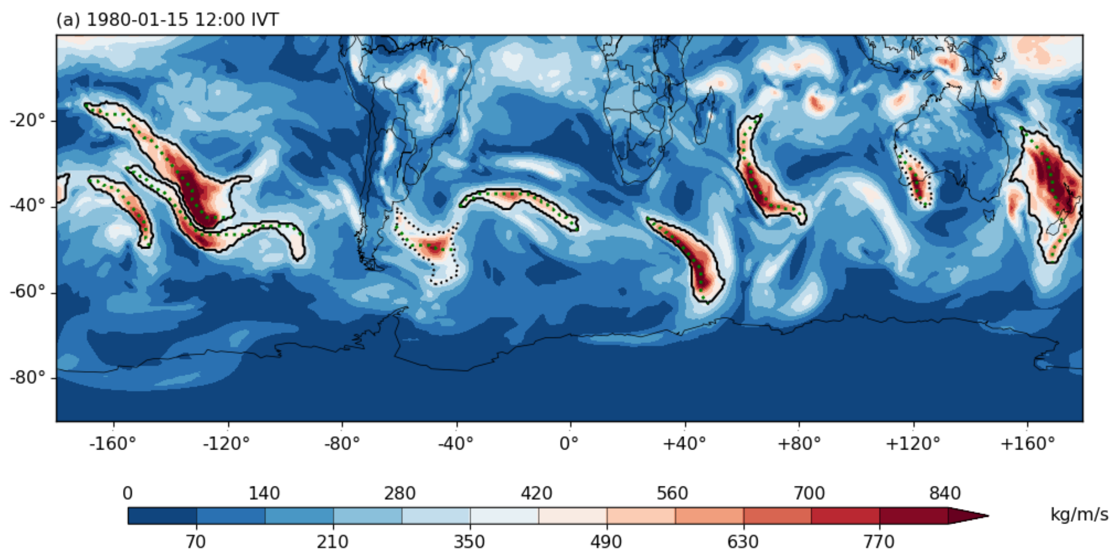

Figure 2: Southern Hemisphere IVT distribution at 1980-01-15 12 UTC. Detected ARs by the THR method are marked out by black contours, the AR axes are drawn as green dotted lines.

Line 160:The AR candidate which is similar to circular region (e.g., when TC occur) will be discarded by this criterion. While, when TC occupies a small partition of the AR candidate (the concurrence of TC and AR), it is hard to be discarded by this criterion.

This is true. Using the IVT information alone it is currently not possible for the method to exclude a TC if it is merged with an AR. One such example is a western Pacific system around 2004-04-14 to 2004-04-17 (Figure 3). Also, when these 2 types of systems are in their transition period it can be rather ambiguous to decide when to classify ARs. In Figure 4 there is one such example. The AR at about $150°W$ is a TC transitioning into an AR, its shape is fairly circular (but still not as circular as to be filtered out). The axis has a big anti-clockwise curve, which quickly unfolds and eventually turns into an AR. We experimented some other metrics, like measuring the ratio of the distance between the 2 end points on the AR axis to the total axis length, if the ratio is too small there is probably too tight a curl in the axis and is more of a TC than an AR. But none of these methods seem to work well to solve this issue. Maybe the relative vorticity could provide some helpful information in such situations, but we can not give any comment on that until enough experiments are done.

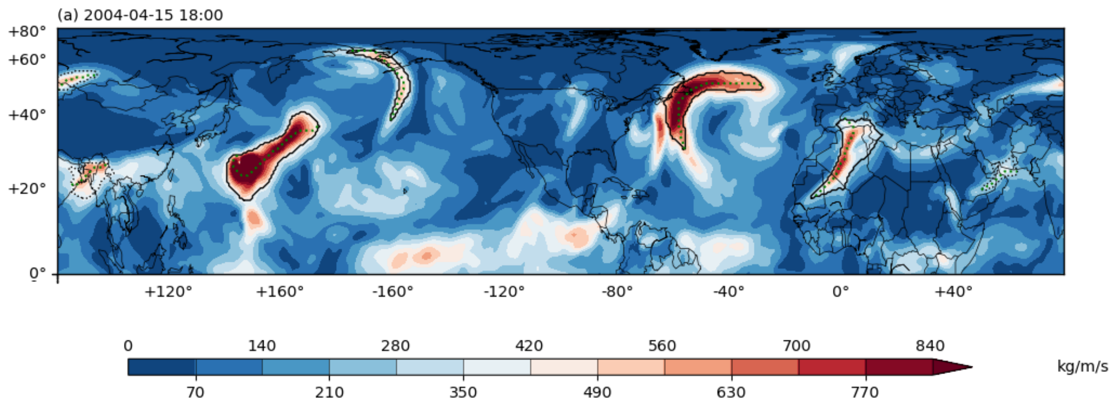

Figure 3: Northern Hemisphere IVT distribution at 2004-04-15 18 UTC. Detected ARs by the THR-t4-s6 method are marked out by black contours, the AR axes are drawn as green dotted lines.

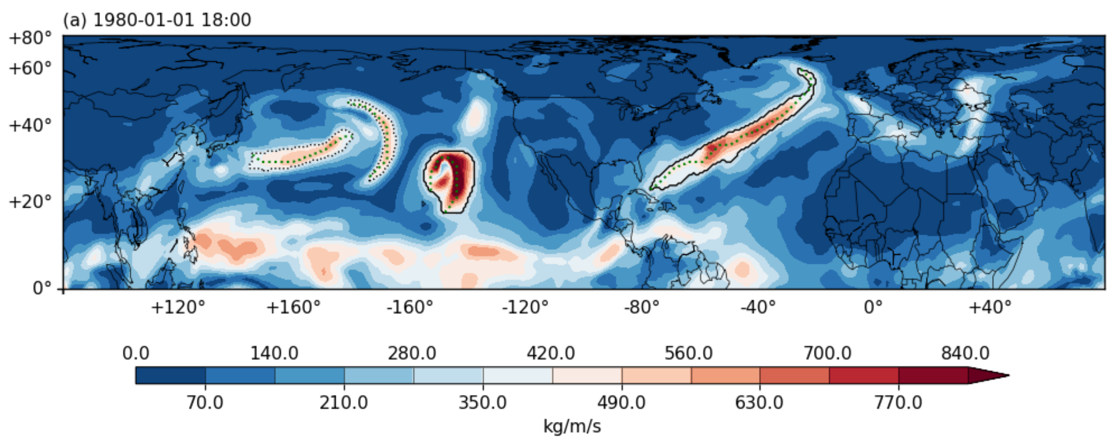

Figure 4: Similar as Figure 3 but for 1980-01-01 18 UTC.

Line 168-169:Whether the tropical moisture need to be considered or not is a controversial topic in the AR study. Usually, the AR is defined as a mid-latitude phenomenon, while the impact of tropical moisture to the mid-latitude climate is not neglectable (Knippertz et al., 2013; Lu et al., 2013). Discarding the ARs whose centroids are within the 23N-23S may discard the genesis stage of an AR event, since the enhanced moisture transport often start from the tropics. So, what kind of tropical AR need to be discarded is deserved to be further considered.

Yes, and the connection to tropical moisture is more obvious in IWV-based AR detections than IVT-based. In the IVT distribution there is a gap created by the transition from tropical trades to mid-latitude westerlies, so the IVT-ARs typically are not as well connected to the tropics as the IWV-ARs. Therefore, the lower latitude bound mostly filters out those westward moving tropical plumes following the trade winds. On the other hand, as you mentioned this is still pretty much in debate, even when an AR as defined by the strong IWV corridor is well connected with the tropical reservoir, how much of the moisture that arrives at the North America continent is actually from the tropics is not quite certain. For instance as discussed in the Bao et al. (2006) paper, they only demonstrated the existence of direct moisture transport along and within such AR corridors, but not the actual percentage contribution. In this study we are taking ARs as mid-latitude systems, which will potentially confine their genesis locations to some extent.

We are adding a bit more to the end of Line 401 to reflect this point:

> Keep in mind that an arbitrary $23°N$ latitude requirement has been applied during the detection stage which is designed to prevent the genesis locations to be traced back to the main moisture reservoir within the tropics.

Section 2.2 geometric considerations.:Several geometric criteria applied here seem arbitrary. The sensitivity test from different geometric criteria should be conducted, especially for the relaxed 800 km length requirement, maximum length and maximum area, isoperimetric quotient and latitudinal range. It will be better to demonstrate how these criteria are selected and what kind of preliminary AR are discarded by these criteria.

These criteria are all determined from the physical natures of ARs (such as the maximum length/area as discussed above), and trial-and-error processes. The proposed method is mostly concerned with relaxing the hard and sensitive IVT strength threshold, so we kept the geometrical considerations largely as controlled variables and did not perturb them in the computations once the trial-and-error stage is finished. As mentioned above, we are still experimenting various ways to better define the AR boundary, including separating the merged ARs. Once the method is regarded as mature, the maximum length/area requirements will not be needed anymore. The dual length requirement, 800 km and 2000 km, is an attempt to introduce some sort of fuzziness into the geometric criteria. It is a hysteresis threshold approach which allows a weaker feature to be retained if and only if in association with a strong feature. More complex fuzzy logic system could potentially be applied, but that goes beyond the current scope of the study.

We are hereby giving some extra results regarding the sensitivities of AR number to minimum length and maximum isoperimetric quotient. These could be included in the Appendix, and point out in the main texts that sensitivity tests have been carried out to arrive at these criteria. The figure below shows the average seasonal AR detection numbers when different (a) minimum length, (b) maximum isoperimetric quotient criteria are used.

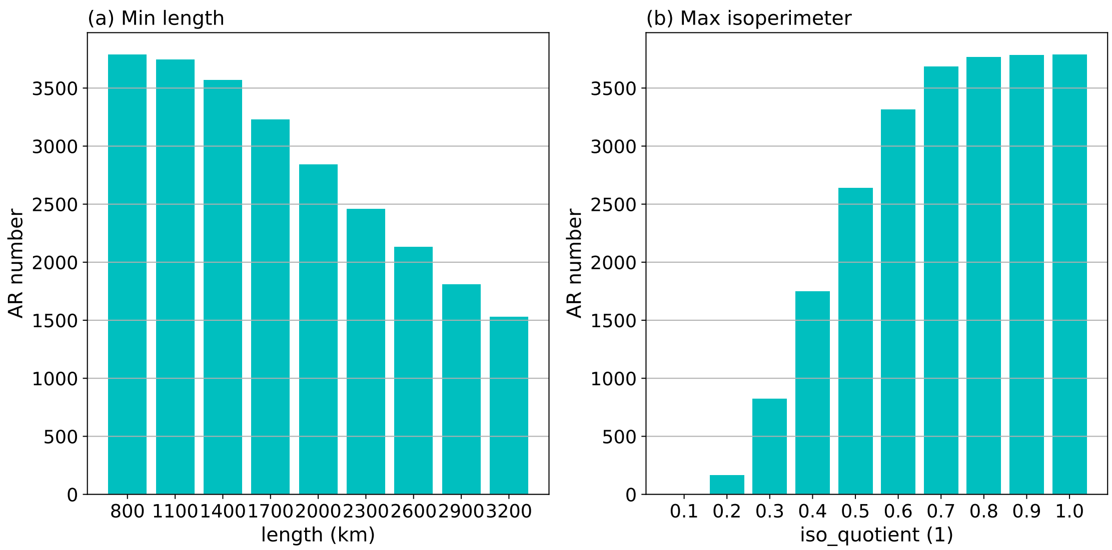

Figure 5: (a) average seaonal (Nov-April) AR numbers in the Northern Hemisphere during 2004-Nov to 2010-April when different minimum length criteria are applied. (b) average seasonal AR numbers in the same time period when different maximum isoperimetric quotient criteria are applied.

Therefore, for minimum length, sensitivity around the 800-1400 km range is fairly small. 2000 km appears to be a standard in the community, and some studies used 1500 km. So 800 km is an even more forgiving criterion. Also keep in mind that the 800 km is used as a relaxed criterion to form a hysteresis thresholding couple with the standard 2000 km criterion, so it does not work quite the same as a single length threshold like in many other studies. In a hysteresis manner, it can avoid the inclusion of noise while still retaining weaker stages of well established ARs.

For maximum isoperimetric, sensitivity around the 0.7 threshold used in this study is also fairly low.

Figure 6 gives some sensitivity tests on the lower latitudinal bound, starting from $11°N$ to $31°N$ with an increment of 2°. Detections north of the this bound are represented in blue color, those south of this bound (but north of $10°N$) are colored in orange. It can be seen that $19 - 23°N$ is the range where the detection number shows reduced sensitivity to the latitude bound choice (Figure 6a). The same is also true for the centroid latitude (Figure 6b) and the latitude of the north-most axis point (Figure 6c). These low-latitude detections have comparable, or even larger lengths than the mid-latitude ones (Figure 6d), but with notably smaller latitudinal span (Figure 6e). Furthermore, such low-latitude detections have primarily westward moisture fluxes, in contrary to the mid-latitude counter-parts (Figure 6f). Therefore, they are mostly zonally orientated, large continuous vapor plumes carried by the tropical trades, and are incompatible with the mid-latitude, storm-related AR definition taken in this study. That is the primary reason for imposing a lower latitudinal bound.

We have explained the choices of maximum length and area above. There are not too many interesting results to show regarding what kind of preliminary ARs are discarded by these criteria. ARs discarded by a given minimum length requirement are, by design, shorter than that; and by a given isoperimetric criteria are, by design, rounder than that. And that is pretty much it.

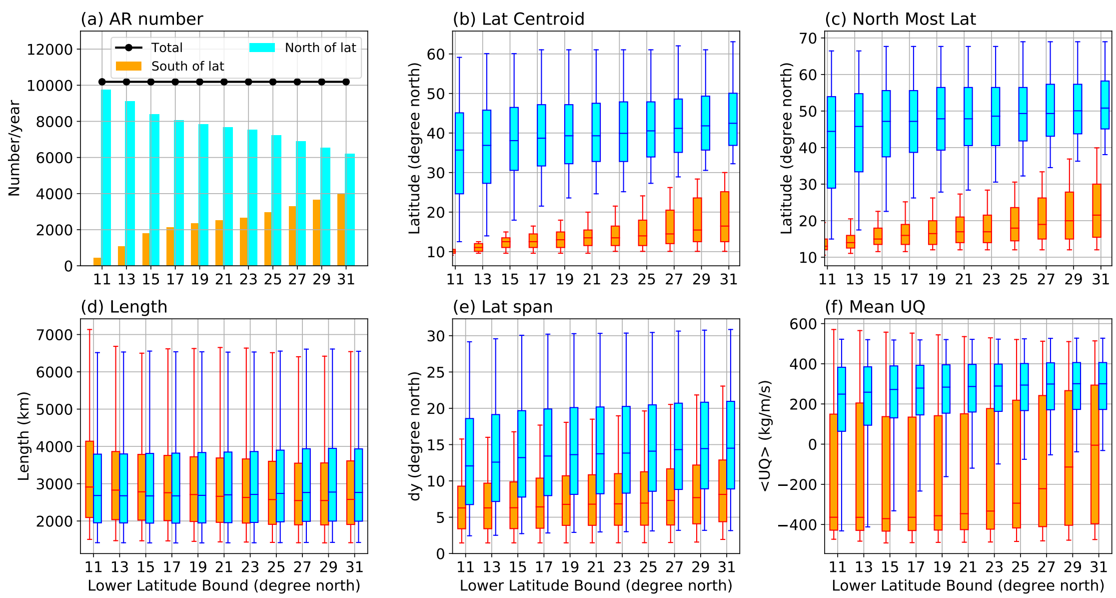

Figure 6: Sensitivity tests on the lower latitudinal bound set to an AR's centroid. (a) average annual AR numbers in the Northern Hemisphere during 2004 to 2008 when different lower latitude bounds are applied. Cyan (orange) bars show the number of ARs whose centroids are north (south) of the given lower latitude bound. Black dotted line is the sum of the two. (b) distribution of the AR latitudinal centroids for ARs north (in blue) or south (in orange) of the given lower latitude bound. (c) similar as (b) but for the distribution of the north-most axis point in ARs. (d) similar as (b) but for the AR lengths. (e) similar as (b) but for the latitudinal span of the ARs. (f) similar as (b) but for the average zonal component of vertically integrated moisture flux (<UQ>, in kg/m/s).

Section 2.3 finding AR axis:This is a very novel method in identifying the AR axis, and it is believed that it has a very good performance in AR with complex shape, orientation and curvature. While, more examples should be provided to demonstrate its good performance in different situations, especially when ARs have varying shape, large curvature, or concurrent with TC or EC.

We could perhaps give a few more plots either in the main text or in supplementary material showing some more AR detections and their axes. This method does give pretty satisfactory results for most of the cases.

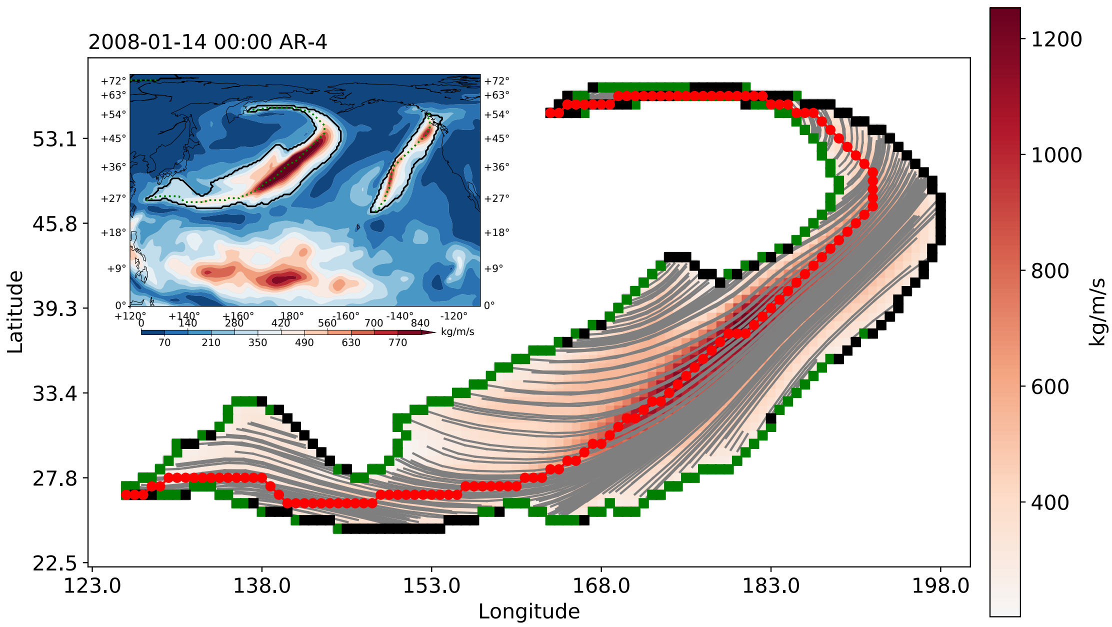

Figure 7: Application of the axis finding algorithm on the AR in the North Pacific, 2008-Jan-14 00 UTC. IVT within the AR is shown as colors, in $kg/m/s$. The IVT streamlines inside the region of the AR are drawn as grey curves. A square marker is drawn at each boundary node, and is filled with green if the boundary node has net input moisture fluxes ($n_i \in L_{k,in}$), and black if it has net output moisture fluxes ($n_i \in L_{k,out}$). The found axis is highlighted in red. The inset image shows the IVT distribution over North Pacific with the detected ARs highlighted in black contour.

In the 2008-01-14 plot (Figure 7), the bigger Pacific AR has a sharp curve at its northern end. I guess a polynomial fitting would have a hard time keeping the fitted curve inside such a narrow band like this. Skeletonization method would work better in this case. The AR to its east has two local maxima (again it is not quite sure whether they should be treated as two), and the found axis follows the maximum IVT value pretty well in this case.

Figure 8 shows a North Pacific AR with a complex shape. After an anti-clockwise bend in the middle it makes a quick clockwise turn at the north end.

Figure 9 is another North Pacific example with a big c-curve in its shape. The IVT directions have a nearly 180 degree rotation following the curvature of the AR. In Guan and Waliser (2015) they imposed a requirement on the coherency of IVT direction and coherency between IVT direction and the object orientation. I am curious how their method would treat this particular AR.

Figure 3 shows an example of coexisting AR and TC. The western Pacific AR has a curly southern tail which is the result of the strong vortex of the TC.

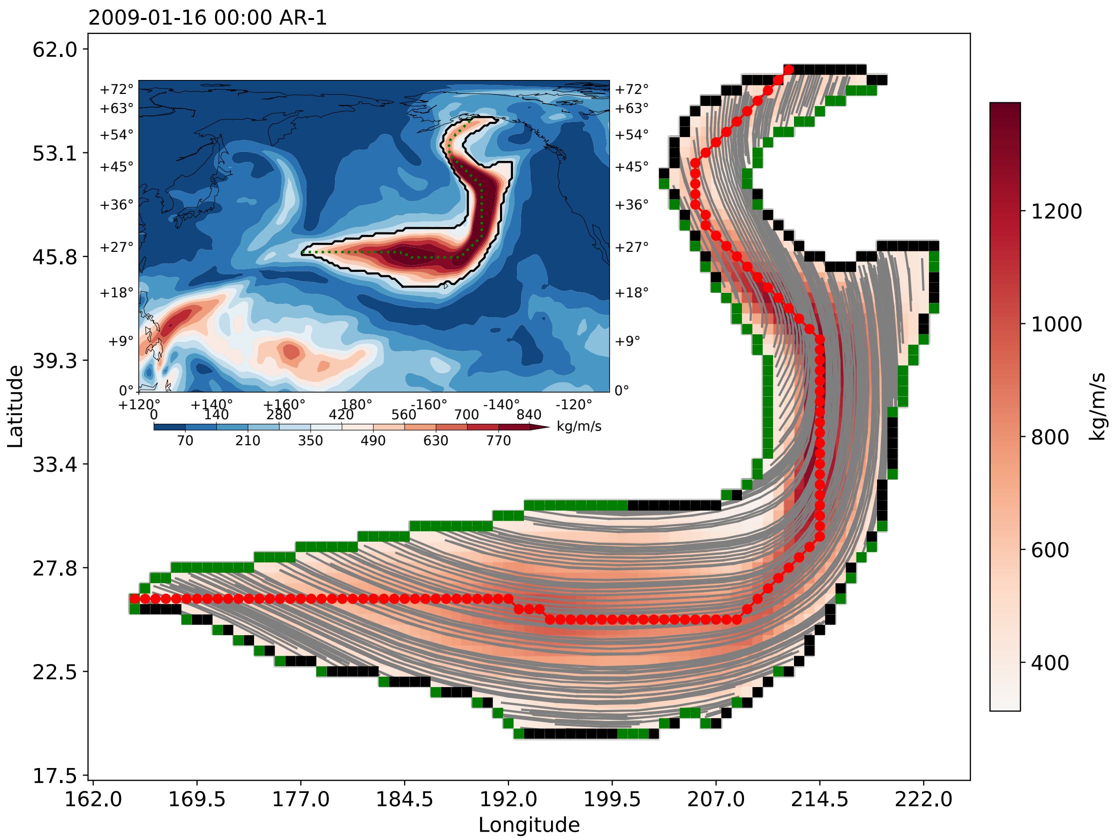

Figure 8: As in Figure 7 but for an AR at 2009-01-16 00 UTC.

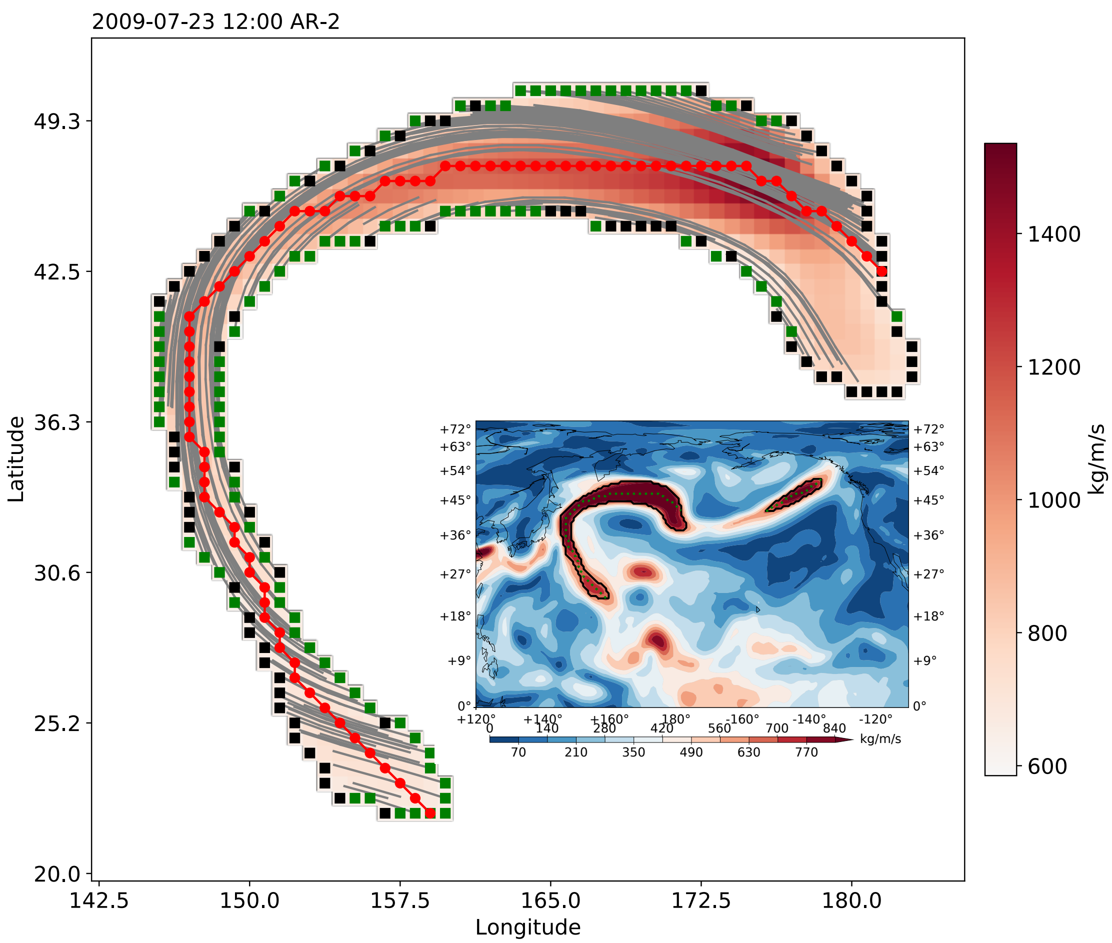

Figure 9: As in Figure 7 but for an AR at 2009-07-23 12 UTC.

In all these three examples, the found axes follow the general flow direction inside the ARs as represented by the streamlines. The directed path design makes sure that it never goes out of the AR boundary. One situation that this method does not work quite well is during a side-to-side merging/splitting event, when there are two local maxima IVT bands in the AR contour, then the algorithm can get confused which one to follow along.

Figure 2: At the end of AR axis, why it no longer exists over the center of AR path way any more (marked by red box in Figure 1)?

The axis does not necessarily follow the center of the AR shape all the time, it is designed to "try" to follow the maximum IVT inside the AR. The reason for it being like this is because this curve has a larger along-axis IVT integral than a path that follows the center of the shape. In fact, this AR is pretty special in that the very north end touches the North Pole, that is why you don't see the few horizontal boundary boxes on the very north end. I guess this strong distortion around here somehow affects the flux direction estimates around this area and subsequently affect the axis finding.

Section 2.4: The Hausdorff distance measurement is an efficient method in assessing the closeness of two AR slices, while the reason why 1200 km is selected as the criterion should be further justified.

The choice of 1200 km as the maximum linkage length is again partially based on existing evidences and partially on trial-and-error processes. (In fact, in this particular case there is not even much of a trial-and-error, the number we threw at it worked so well we did not bother testing more after a few perturbed numbers around 1200 km.) We are again making an implicit association between ARs with extra-tropical cyclones. In the Neu et al. (2013) work they listed a number of extra-tropical cyclone tracking methods, and many of which use a maximum distance as the nearest neighbor requirement similar as in this study. Choices of 600 km, 800 km and 1000 km are included in their supplementary table A (for 6-hourly intervals). We picked the largest one and gave it an extra margin to make it 1200 km, because in addition to movement, length variations also contribute to Hausdorff distance. It turns out that the choice of this length has very low sensitivity, I would say much lower than a minimum areal overlap ratio when linking two consecutive ARs based on their areal overlaps. Because when the AR slides sideways, the areal overlap will drop very quickly, and it is even more so when the data have a coarse temporal resolution. In this Hausdorff distance method, setting the number relatively higher won't affect the result, because it always chooses the closest pair. Setting it too small would tend to break the track, but if one chooses a sensible value depending on the data resolution, it will be pretty robust. We have tested 3-hourly, 6-hourly and daily interval data and it all worked quite well.

I also feel that we are sometimes using the critics of "arbitrary" too lightly. If arbitrariness is the most important question to ask, it may be asked way too late. How are the magic 250 kg/m/s (2 cm for IWV) and 2000 km numbers arrived at? How do they physically related to this kind of phenomenon? Has anyone been bothered to do a scale analysis to provide some physical support to these magic numbers? To me, the choice of 250 kg/m/s IVT threshold appears equally arbitrary, the 2000 km length requirement that so many studies have been following seems equally arbitrary, the zonal maximum and average based formula used by one of the pioneering works of Zhu and Newell (1998) seems equally arbitrary, and the 85th percentile+100 kg/m/s combination is double-arbitrary. However, I should have replaced every "arbitrary" with "reasonable", because whatever number they came up with, they must have arrived at the number very carefully, after iterations and iterations of trial-and-error, making various compromises to achieve the best fit for the dataset/datasets they experimented with.

This also made me think why when we need to make a yes-no decision using numbers, a simple threshold is almost always the first solution that comes into mind? Even in some of the very complicated algorithms, for instance some of the fuzzy-connected image segmentation algorithms that see applications in medical image processing, after some complex computations of fuzzy adjacency, affinity and connectedness, people use a simple threshold on the connected map to obtain the final object. The entire class of level-set methods is based on single number thresholding (except that they always put the threshold at 0). (Yes, we actually experimented both of them, although not very successfully.) Even in some neural networks that claim to mimic the functioning of neural systems, we sometimes see that after layers of layers neuron propagation, people take a simple thresholding on the final activation function output to make a binary classification. It just seems that a clear-cut thresholding is such an intuitive mindset that it makes it difficult for us to think alternatively.

Considering the concerns regarding such geometric criteria choices, we are proposing a dedicated section in the Appendix to explain them:

A5 Explanations for the choices of the geometrical filtering criteria

The geometrical filtering criteria, including the maximum length/area, the minimum length, the maximum isoperimetric quotient, the minimum centroid latitude, and the maximum Hausdorff distance in the nearest neighbor linkage process, are all determined from the physical natures of ARs as well as trial-and-error processes. The proposed method is mostly concerned with relaxing the hard and sensitive IVT strength threshold, therefore the geometrical considerations are largely treated as controlled variables. Some further details regarding the choices of these criteria are given here.

- Maximum length/area

  The maximum area/length requirements are set to fairly large values. The maximum length is set to 11,000 km, which is longer than the great circle distance (through Pacific, $\sim$ 10400 km) from Hong Kong ($\sim 22.3°N, 114.2°E$) to Seattle ($\sim 47.6°N, -122.3°W$). Assuming an average width of 700 km (which is on the larger side for typical ARs), this multiplies to an area of $\sim 7.3 \times 10^6 km^2$, which is smaller than the maximum area requirement set here ($10 \times 10^6 km^2$). These requirements are far larger than what could be expected from real world AR sizes, therefore they would only filter out erroneous detections.

  [Insert Figure 1 here.]

  Regions with such large sizes only happen when two or more ARs are grouped together in one contour, or some of the ARs are connected with the large and continuous moisture plumes in the tropics to form a large continuous region. This happens more frequently when the structuring element E is set too large, as explained in Section 3.1. Figure 1 gives one such example. In Figure 1a, the structuring element is set to t7-s6. This merges the Pacific ARs with tropical plumes to form a big connected region that later gets filtered out. Figure 1b is using the recommended t4-s6 parameters, this time the mid-latitude signals are better separated from tropical ones so the Pacific AR is correctly retained. Those small noisy contours will be filtered by the minimum area requirement.

  A method is being developed to separate ARs that have been mistakenly grouped together, once the new algorithm is fully ready, such maximum length/area criteria are no longer needed.

- Minimum length and maximum isoperimetric quotient

  Some results regarding the sensitivities of AR numbers to the minimum length and maximum isoperimetric quotient criteria are given in Figure 5.

  [Insert Figure 5 here.]

  For minimum length (Figure 5a), sensitivity around the 800-1400 km range is fairly small. Also keep in mind that the 800 km threshold is used as a relaxed criterion to form a hysteresis thresholding couple with the standard 2000 km criterion, so it does not work quite the same as a single length threshold like in many other studies. This is an attempt to introduce some fuzziness into the geometric criteria, and it allows a weaker feature (above 800 km in length) to be retained if and only if in association with a strong feature (longer than 2000 km).

  For maximum isoperimetric, sensitivity around the 0.7 threshold used in this study is also fairly low (Figure 5b).

- Minimum centroid latitude

  [Insert Figure 6 here.]

  Figure 6 gives some sensitivity tests regarding the choice of the lower latitudinal bound imposed on AR region's centroid. After detecting all ARs north of $10°N$ during 2004-2008, they are separated into two groups depending on their centroid being north or south of a given lower latitudinal bound. Values of this bound range from $11°N$ to $31°N$ with an increment of $2°$. Detections north of the this bound are represented in blue color in Figure 6, those south of this bound (but north of $10°N$) are colored in orange.

  It can be seen that $\sim 19 - 23°N$ is the range where the detection number shows reduced sensitivity to the latitudinal bound (Figure 6a). The same is also true for the centroid latitude (Figure 6b) and the latitude of the north-most axis point (Figure 6c). These low-latitude detections have comparable, or even larger lengths than the mid-latitude ones (Figure 6d), but with notably smaller latitudinal span (Figure 6e). Furthermore, such low-latitude detections have primarily westward moisture fluxes, in contrary to the mid-latitude counter-parts (Figure 6f). Therefore, they are mostly zonally orientated, large continuous vapor plumes carried by the tropical trades, and are incompatible with the mid-latitude, storm-related AR definition taken in this study. This is also the primary reason for imposing a lower latitudinal bound.

- Maximum Hausdorff distance

  The choice of 1200 km as the maximum Hausdorff distance during the track stage is based on references to similar maximum distance requirements used in extra-tropical cyclone tracking practices (Neu et al. (2013)), and trial-and-error processes. Choices of 600 km, 800 km and 1000 km (for 6-hourly intervals) are included in the supplementary Table A of (Neu et al. (2013)). We chose the largest one and gave it an extra margin to make it 1200 km, because in addition to movement, length variations also contribute to Hausdorff distance. The choice of this length has very low sensitivity, and can be easily adjusted for data with different temporal resolution (e.g. scaled to 600 km for 3-hourly data, or 2400 km for 12-hourly).

Line 278-283: From the histogram of occurrence and the box plot of mean IVT (Figure5e), we can observe that the IVT85% method detect more ARs but with lower IVT intensity. From my experience, the reason would be that the 100 kg/m/s is too low to detect a lot of weak ARs near the polar region. So, in the comparison of different detection methods (THR, constant IVT thresholding method, and percentile method), the AR frequency maps are suggested to be provided, which can help to further explore the difference of spatial distribution among the ARs detected by different methods, and the region preference of different methods.

Yes, we also observed that the IVT85% method gives more AR occurrences at higher latitudes than the other 2 methods. We believe this is mostly because of the 85th percentile threshold design, and we are not experienced enough with this method to comment on the fixed 100 kg/m/s threshold in the pole regions. We did not include such spatial distribution maps mostly because we intended to make this work primarily focused on methodology description. Inter-method comparisons are discussed in much detail in a separate work (now published at atmosphere (doi: 10.3390/atmos11060628). Some IVT85% and IVT250ano results are put here (with no duplicates with the atmosphere paper) is because to demonstrate the reduced parameter sensitivity in the proposed method, we have to give it something to compare against. Other than that inter-method comparison has been reduced as much as possible.

Figure 5: To be honest, it is hard to justify which result is better for the two examples marked by red box in Figure 2, THR or IVT85%. The ARs detected by THR (green contour) have two peaks, and the IVT85% divide them into two ARs (black dash contour). We can not say that the result of THR is better. Maybe the two peaks are controlled by different systems and should be diagnosed separately.

It is not our intent to (and we did not) make any claim in the texts that the THR results are better in these cases. On the contrary, we stated in Line 330-334 that such ambiguities in the THR results contribute to geometry-related uncertainties. As mentioned above (and also in the text in Line 330-331, and Line 450-454), we are working on some methods to separate such merged systems, but they are not implemented in this study. That is why we decided to put these two cases there to demonstrate the existing problems of the method rather than cherry-picking some other time slices that clearly show better results in the THR results.

Line 344-348: How to prove that the low IVT mean for large size ARs is not due to the inclusion of weaker systems?

The deduction we followed is as such: the mean IVT is lower in larger sized ARs. This could be a result of (i) AR sizes are larger so the mean get more diluted; or (ii) the IVT values are lower (inclusion of weaker systems) to begin with; or a combination of (i) and (ii). If (ii) is the primary contributor, the maximum IVT would also be low, because that does not get averaged as in the computation of mean IVT. But we observed much smaller differences in maximum IVT than in mean IVT, therefore (ii) is not the primary reason, therefore the size dilution is the main reason for lower mean IVT.

Minor comments:

Line 60-65: Another AR axis identification method is proposed in (Pan & Lu, 2019), which can fit smooth curve for ARs with varying shape, orientation and curvature. It will be a good reference here.

Thanks for the information. We could add it here.

Line 247: The "constant IVT threshold approach" should be adjusted to the "constant IVT anomaly threshold approach" to distinguish from the real constant IVT threshold approachs (e.g., Sellars et al., 2017).

That is a fair point. We are making such a change.

References

J-W. Bao, S. A. Michelson, P. J. Neiman, F. M. Ralph, and J. M. Wilczak. Interpretation of enhanced integrated water vapor bands associated with extratropical cyclones: Their formation and connection to tropical moisture. 134:1063–1080, 2006. ISSN 0027-06441520-0493. doi: 10.1175/mwr3123.1. URL `http://dx.doi.org/10.1175/mwr3123.1`.

Bin Guan and Duane E. Waliser. Detection of atmospheric rivers: Evaluation and application of an algorithm for global studies. *Journal of Geophysical Research: Atmospheres*, 120(24):12514–12535, 2015. doi: 10.1002/2015JD024257.

Urs Neu, Mirseid G. Akperov, Nina Bellenbaum, Rasmus Benestad, Richard Blender, Rodrigo Caballero, Angela Cocozza, Helen F. Dacre, Yang Feng, Klaus Fraedrich, Jens Grieger, Sergey Gulev, John Hanley, Tim Hewson, Masaru Inatsu, Kevin Keay, Sarah F. Kew, Ina Kindem, Gregor C. Leckebusch, Margarida L R Liberato, Piero Lionello, Igor I. Mokhov, Joaquim G. Pinto, Christoph C. Raible, Marco Reale, Irina Rudeva, Mareike Schuster, Ian Simmonds, Mark R. Sinclair, Michael Sprenger, Natalia D. Tilinina, Isabel F. Trigo, Sven Ulbrich, Uwe Ulbrich, Xiaolan L. Wang, and Heini Wernli. Imilast: A community effort to intercompare extratropical cyclone detection and tracking algorithms. 94:529–547, 2013. ISSN 00030007. doi: 10.1175/BAMS-D-11-00154.1.

Yong Zhu and Reginald E. Newell. A proposed algorithm for moisture fluxes from atmospheric rivers. *Monthly Weather Review*, 126:725–735, 1998. ISSN 0027-0644. doi: 10.1175/1520-0493(1998)126<0725:APAFMF>2.0.CO;2.