# Referee Comment #2

**General comments:**

1. The relationship between the dust, sea salt and the AR mechanisms needs to be more clearly and directly shown. The thickness diagnostics (and differences) for the composite “common” ARs are hard to interpret. I would recommend keeping these diagnostics for the case studies, but illustrating the connection between the ARs and aerosols cluster groups more directly, or in a more focused way. Ideas include using AR variables themselves,(IVT, IWV, or low level winds) for the “strengthening/weakening” component with the aerosols in lat/lon space, rather than box/whisker, and only showing the clusters that are significant. Or perhaps AR-spine centric averages vs aerosols cluster (highest density areas?)(and/or perhaps thickness) in scatter plots, to show this relationship. I think it is there, but at present it is a little unfocused. Also, significance needs to be shown in any difference plots.

Thank you for your suggestions. We have substantially improved the physical description of the processes that relate aerosols and the observed AR differences in the revised manuscript, especially for the ARI-BASE comparison, thanks to some comments of Referee #1.

In an early stage of our research, we also considered the analysis/representation of the IVT or IWV fields instead of the thickness diagnostics to analyse the clusters. We have a collection with all the ARs IVT representations available. However, as we have discussed later in the Specific Comments section, representing e.g. the IVT fields of all the members of a cluster (or averaging them) made quite difficult the extraction of any conclusions. The natural variability of ARs, with their diverse trajectories, locations, width, etc. obscured/hided the patterns in the differences of the group, thus complicating the relation with aerosols effects. After trying many approaches, we came up with the present methodology. Furthermore, our physical discussion is mainly based on temperature changes due to aerosols effects, and the thickness fields show these changes. However, thanks to some of your specific suggestions, we have added the trajectories of each AR belonging to a cluster and their mean trajectory to the thickness plots of the most relevant clusters (Fig. 9 and 12). For instance, you can find the resulting plot of the ARI clusters 2-3 in the answer to the "case studies" specific comment. Each thin arrow represents an AR. It is located on its mean latitude with its mean direction and the length of the arrow is proportional to its mean intensity (IVT). The thicker arrow represents the mean characteristics of the ARs belonging to the cluster. We sincerely hope that the revisions we have implemented address your concerns.

In response to your comment regarding the significance of the differences, we will include statistical significance to the greatest extent possible, where it is applicable. Given the small sample size of some clusters, we are aware that statistical analyses may have limitations. In cases like this, it may be more reasonable to focus on providing a qualitative description of the observed differences. That being said, we would like to once again express our gratitude for your valuable feedback.

2. AR and ARDT uncertainty needs to be addressed. AIRA needs to be put into context of published ARDTs, and specifically, regional-specific algorithms that cover the Iberian Peninsula (e.g. IDL/Ramos, Lavers, Brands). Given the IDL code uses transects and also a Lagrarangian framework, this is the most similar type of code). ARTMIP (https://www.cgd.ucar.edu/projects/artmip/algorithms will have the reference list for the above mentioned ARDTs) has robustly shown that threshold choice is the largest source of AR metrics variability across ARDTs with dramatic differences in frequency, for example, depending on how this is chosen. See specific comments for details on suggestions on how to address this issue.

We strongly appreciate your comments and suggestions here. It is something that was missing and the other referees also noted this issue. In the revised manuscript, we are going to put AIRA in the ARTMIP context and classification, including its main differences with the IDL/Ramos, Lavers and Brands ARDT algorithms, which are the most similar to AIRA and also detect ARs over the Iberian Peninsula. As a preliminary observation, the main contrast is that these algorithms make use of spatial tracking, while AIRA never uses it, as it is intended to perform also in regions close to the domain edges. This is indeed the case in our study, with the detection lines located very near the limits of the spatial domain.

With respect to the AR and ARDT uncertainty, we have re-ran AIRA multiple times with different IVT and duration thresholds to assess the sensibility to the thresholds choice. For information about the results, we would like to refer to the Specific Comments section.

3. Referencing needs to be improved and representative of the recent AR literature.

Thank you for your remark. We have substantially improved the referencing of this work in the revised manuscript thanks to not only your valuable suggestions but also the recommendations of the other two referees.

**Specific comments:**

Line 21: In the midlatitudes, this is indeed the case, but not necessarily for high latitude ARs. I recommend amending this statement with “in the midlatitudes”.

Thank you very much for your comment. We have corrected it.

Lines 24 and 25: There are many many references that could fit this statement, I recommend adding an “e.g.,” to your citation list, or add a few more references.

Thank you for the remark. Many other references could have been used here, so we have added "e.g." to the revised manuscript, as the included references were just some examples of researches about ARs in those regions.

Lines 28,32,34: Again, there are quite a few that could be listed here, so “e.g.” should be used. I am surprised not to see any Lavers references as this group was among the first to discuss North Atlantic ARs.

We appreciate again your remark. We have corrected it by using "e.g." and we have also included a Lavers reference.

Paragraph Line 36: I appreciate the author's discussion here, but there are some major gaps in the literature review. ARTMIP has had a number of workshops, plus 5 major group/overview papers, and many contributed papers. All discuss the issues of defining and detecting ARs, and the philosophy of using an ARDT (AR detection tool) that is appropriate for the science question asked. In addition to referencing the workshop report (or instead of), please read and cite the following papers. (Note: the climate change papers, O’Brien and Shields/Payne, would be good additions to the climate change literature review sentences, with the Rutz and Collow papers for reanalysis).

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., Ullrich, P., Gershunov, A., Goldenson, N., Guan, B., Qian, Y., Ramos, A. M., Sarangi, C., Sellars, S., Gorodetskaya, I., Kashinath, K., Kurlin, V., Mahoney, K., Muszynski, G., Pierce, R., Subramanian, A. C., Tome, R., Waliser, D., Walton, D., Wick, G., Wilson, A., Lavers, D., Prabhat, Collow, A., Krishnan, H., Magnusdottir, G., and Nguyen, P.: Atmospheric River Tracking Method Intercomparison Project (ARTMIP): project goals and experimental design, Geosci. Model Dev., 11, 2455-2474, https://doi.org/10.5194/gmd-11-2455-2018, 2018.

Rutz, J.J, Shields, C.A., Lora, J.M, Payne, A.E., Guan, B., Ullrich, P., O'Brien, T., Leung, L.-Y., Ralph, F.M., Wehner, M., Brands, S., Collow, A., Goldenson, N., Gorodetskaya, I., Griffith, H., Hagos, S., Kashinath, K., Kawzenuk, B., Krishnan, H., Kurlin, V., Lavers, D., Magnusdottir, G., Mahoney, K., McClenny, E., Muszynski, G., Nguyen, P.D., Prabhat, Qian, Y., Ramos, A.M., Sarangi, C., Sellars, S., Shulgina, T., Tome, R., Waliser, D., Walton, D., Wick, G., Wilson, A., Viale, M.: The Atmospheric River Tracking Method Intercomparison Project (ARTMIP): Quantifying Uncertainties in Atmospheric River Climatology, Journal of Geophysical Research-Atmospheres , https://doi.org/10.1029/2019JD030936, 2019.

O’Brien, Travis Allen and Wehner, Michael F and Payne, Ashley E. and Shields, Christine A and Rutz, Jonathan J. and Leung, L. Ruby and Ralph, F. Martin and Marquardt Collow, Allison B. and Guan, Bin and Lora, Juan Manuel and et al., (2022) Increases in Future AR Count and Size: Overview of the ARTMIP Tier 2 CMIP5/6 Experiment. JGR-A https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2021JD036013.

Collow, A.B., Shields, C.A., Guan, B., Kim, S., Lora, J.M., McClenny, E.E., Nardi, K., Payne, A., Reid, K., Shearer, E. J. , Tome, R., Wille, J.D., Ramos, A.M., Gorodetskaya, I.V., Leung, L.R., O’Brien, T.A., Ralph, F.M., Rutz, J. Ullirich, P.A., Wehner, M., (2022) An Overview of ARTMIP’s Tier 2 Reanalysis Intercomparison: Uncertainty in the Detection of Atmospheric Rivers and their Associated Precipitation, Journal of Geophysical Research, Atmospheres, https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2021JD036155.

Shields, C. A., Payne, A. E., Shearer, E. J., Wehner, M. F., O’Brien, T. A., Rutz, J. J., Leung, L.R., Ralph, F. M., Collow, A. B. M., Ullrich, P. A. Ullrich, Dong, Q., Gershunov, A., Griffith, H., Guan, B., Lora, J. M., Lu, M., McClenny, E., Nardi, K. M., Pan, M., Qian, Y., Ramos, A. M. Ramos, Shulgina, T., Viale, M., Sarangi, C., Tomé, R., Zarzycki, C. (2023). Future atmospheric rivers and impacts on precipitation: Overview of the ARTMIP Tier 2 high-resolution global warming experiment. Geophysical Research Letters, 50, e2022GL102091. https://doi.org/10.1029/2022GL102091

More details on ARTMIP here: https://www.cgd.ucar.edu/projects/artmip

We kindly appreciate all your suggestions and recommended references. We have improved and extended this discussion in the revised manuscript. Furthermore, some of these references were also good additions to other parts of the text, as you just mentioned.

Line 48: The statement that GCM's "may not accurately represent their (AR) behavior" is a bit misleading. Most GCMs (and ESMs) are able to simulate the synoptics, bulk numbers, duration, etc. realistically. I recommend amending this statement specifically to AR-precipitation, given it is the precipitation piece that does better with high resolution (citations are needed here, there are quite a few out there now for high resolution global/earth system models, and ARs).

Thank you for your comment. We have amended it in the revised manuscript.

Line 51: I am not sure I understand why a timeslice approach doesn't work for limited area models? Many timeslice ARDTs work well within a limited area domain (see the ARDT list on the ARTMIP webpage, some of these are both timeslice and regional). I agree with the authors that regional ARDTs tend to do a better job because localized considerations are made for regional-specific that would not otherwise be considered in globals (for example, for IP, the complex topography and the North Atlantic storm track climatology). If this is the intent of the authors, I recommend using this as motivation for the newly developed ARDT for the IP, rather than timeslice vs lagrangian approach. If I misunderstood, please make this statement more clear.

Many ARDTs work well within a limited area domain (regional domain) if it is big enough to perform the spatial tracking (mainly over the ocean) usually required to determine the length of the AR. In our case, the detection lines were very close to the limits of the study domain, as you can see on Figure 1 (red box, inner domain). Therefore, we introduced a duration-length relation to estimate the length of the AR, allowing us to work with smaller regions and thus reducing the time to perform computationally costly simulations such as online aerosol runs to understand ARs mechanisms. As referee #3 commented, the innovation of AIRA relies on overcoming the RCMs limitations where most of the runs are focused over land, and this precludes capturing the long way over the ocean. This was the motivation to develop this new regional ARDT, not only the higher resolution (which is also an advantage that plays an important role in the study of AR behaviour and AR-related precipitation at the local scale). The statement in line 51 has been corrected, specifying the cases in which spatial tracking given a fixed time step method is not suitable (not enough domain to perform the tracking). We are going to include this motivation as clear as possible in the revised manuscript and we want to thank you again for the interesting questions.

Introduction general comment: I am surprised there is no mention of the Calwater experiment. Although this was focused on the western U.S., it was an important and groundbreaking study to look at aerosols with observations and AR. Here is a citation from CalWater that uses the same model as this study, i.e. WRF-Chem.

Naeger, A. R. (2018). Impact of dust aerosols on precipitation associated with atmospheric rivers using WRF-Chem simulations. Results in Physics, 10, 217-221, https://www.sciencedirect.com/science/article/pii/S2211379717318223

Thank you very much for this comment, we have added a brief mention to this study in the Introduction section of the revised manuscript.

Paragraph at line 74: It might be useful to readers familiar with climate models, but not WRF forecast systems, to add a sentence or two explaining how lateral boundary conditions nudge the model back to the “observations”. This is important for when you describe your common ARs periods later, it makes sense to use common periods given each simulation is reproducing the same forecast period, but just with different aerosol treatments. If I am misunderstanding the design, please clarify.

Thank you for your comment. Other referees have requested a brief explanation about the model set up and a more profound explanation about the experiments. Although the complete description of the three simulations can be found in the reference included in line 85 (Jerez, S., Palacios-Peña, L., Gutiérrez, C., Jiménez-Guerrero, P., López-Romero, J. M., Pravia-Sarabia, E., and Montávez, J. P.: Sensitivity of surface solar radiation to aerosol–radiation and aerosol–cloud interactions over Europe in WRFv3.6.1 climatic runs with fully interactive aerosols, Geoscientific Model Development, 14, 1533–1551, https://doi.org/10.5194/gmd-14-1533-2021, 2021), we are going to include a brief description in the revised manuscript.

In answer to your question, boundary conditions from the GCM were updated every 6 h to the outer domain. Although nudging was applied to the outer domain, neither nudging nor re-initialization of initial conditions have been used in the target (inner) domain. We were interested in allowing the model to run "freely" in this domain once the initial conditions had been stablished, in order to see how the different aerosol treatments affected the simulations. We will address these comments in the brief explanation of the experiments in the revised manuscript.

Line 88: Just checking how “online” is meant here, as an active coupled component and not stand-alone simulation?

Thank you for your question. Yes, that's exactly what is meant here. We have added a little explanation to that sentence: "In the ARI experiment, aerosols were treated online, introduced as an active fully coupled component, and the aerosol-radiation interactions were activated in the model".

Line 108: I think this a Lagrangian approach, i.e. tracking rather than timeslice, given Figure 2? I am not sure I understand why a regional ARDT can’t track an AR? This approach is similar to the IDL ARDT (an ARTMIP contributor, Ramos et al., 2016). I think it would be helpful to add what aspects of AR science that AIRA addresses that the IDL does not. Or, how it compares to IDL, especially given both of these ARDT look at Iberian ARs.

Ramos, A. M., Nieto, R., Tomé, R., Gimeno, L., Trigo, R. M., Liberato, M. L. R., and Lavers, D. A.: Atmospheric rivers moisture sources from a Lagrangian perspective, Earth Syst. Dynam., 7, 371–384, https://doi.org/10.5194/esd-7-371-2016, 2016

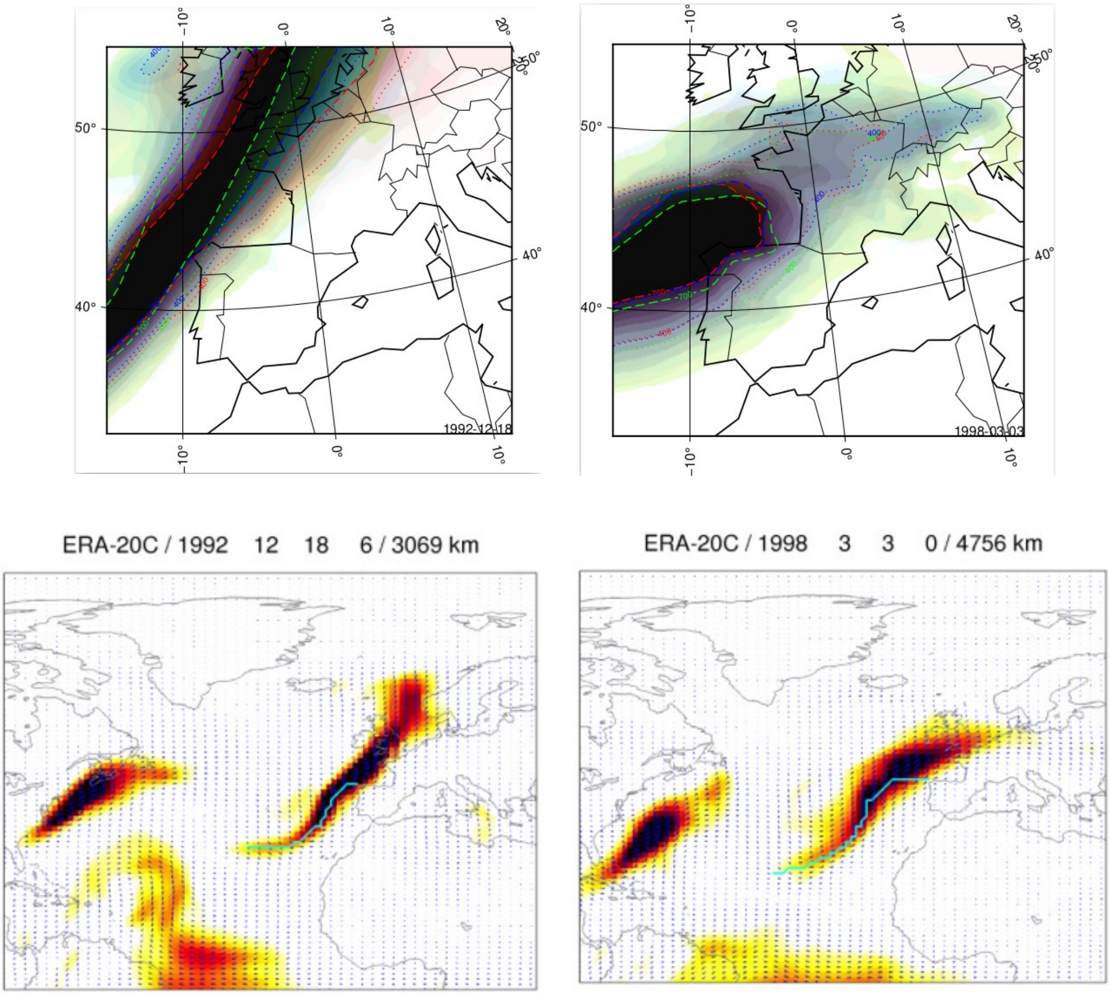
Thank you for your question. AIRA never uses spatial tracking, because the detection lines are so close to the domain limits that it would not be possible to do it. This is the main difference with the IDL Ramos approach, because it performs the tracking to estimate the length of the AR, but we have introduced a duration-length correspondence (given an estimation of the wind speed of ARs in the studied area). However, we are going to include a comparison between the IDL ARDT and Brands ARDT with respect to AIRA in the revised manuscript, as suggested by your second general comment and by the other referees. With respect to the second question here, ARDTs can track ARs if the regional area it is working on is wide enough to perform the spatial tracking. It was not the case of our region. We would like to refer to the answer given to the specific comment about line 51.

Line 134 and Paragraph at Line 185: From Table 1 and paragraph at Line 185, I think this is an absolute threshold, used for all simulations and does not change with the respective simulated climatologies? If so, please state that an absolute threshold is used for all simulations in the initial description, and point to the application for further explanation.

You are right, it is an absolute threshold stablished by the user and we have chosen to use the same value for all three simulations. Following your suggestion, we will state in this section that it is an absolute threshold used for all simulations, and we will refer to the AIRA implementation section for more information.

Line 196: Which ARDT catalogues/datasets were compared? The Brands ARDT contributions to ARTMIP are regional algorithms.

As answered to Referee #3, by the time this research was conducted, there was a website mentioned in Brands et al. (2017) with their *Atmospheric Rivers Archive* available: http://www.meteo.unican.es/atmospheric-rivers. This catalogue documented all the ARs detected by their algorithm using ERA-20C data and we compared our results with it (see figure below for some qualitative examples). Unfortunately, the page was shutdown. To answer Referee #3 questions, we have contacted the authors and they have provided us a database with all the information through a Zenodo repository (https://doi.org/10.5281/zenodo.8010794), although the representation tool is not available anymore. In the revised manuscript, we are going to assess the coincidences to the fullest extent possible.



1992-12-18
1998-03-03
ERA-20C / 1992 12 18 6 / 3069 km
ERA-20C / 1998 3 3 0 / 4756 km

Figure: ARs identified the 1992-12-18 (left) and 1998-03-03 (right) by AIRA (top) and Brands ARDT (bottom, Brands et al., 2017). In the top images, green, red and blue contours/shades represent the ARs of the BASE, ARI and ARCI simulations, respectively.

Line 203: This is consistent with ARTMIP findings as October being the month with the maximum frequency for these latitudes (Rutz et al. 2019, Fig 13).

Thank you very much for your remark. We have added this citation to the manuscript:

"Notably, the highest number of ARs is detected in October, with at least 30 ARs identified in all three simulations (Figure 4 (top)). This result is consistent with the findings of Rutz et al. (2019)"

Line 207: The mean intensity values are somewhat “baked in” to the values given the application of an absolute threshold.

Thank you for your comment. The lower the IVT threshold, the lower the mean intensity of the identified ARs, and viceversa. However, to identify ARs we have to set an IVT threshold, either absolute or relative. For precise information about how the mean intensity of the ARs in each simulation changes with the IVT threshold, see the answer to the last question, where we try to assess the variability of AIRA.

Figure 5: I noticed is that the AR metrics presented in this paper do not agree with other published results that look at aerosols, ARs, and climate, (Baek et al., 2021)

where the Baek shows very little change over the Iberian Peninsula in the thermodynamic/precipitation and more of a change with the dynamics. There could be many reasons, including model resolution, aerosol treatment, ARDT, but this should be discussed or addressed in some way.

Baek, S.H., Lora, J.M. Counterbalancing influences of aerosols and greenhouse gases on atmospheric rivers. Nat. Clim. Chang. 11, 958–965 (2021). https://doi-org.cuucar.idm.oclc.org/10.1038/s41558-021-01166-8

Thank you for your comment. Referee #3 has suggested mentioning this paper in the Introduction section, although it seems relevant to include it also during the AR-related precipitation discussion. As a preliminary comparison, our approach in Fig. 5 is similar to Extended Data Fig. 2 (% AR Precip Relative to Total Precip) of said paper, and we even use the same metrics (a percentage). For the historical period of their study (1920-2005), the authors have obtained an AR-related precipitation between 20 and 40% of the total accumulated precipitation over the North Athlantic coast of the IP. These results are consistent with our outcomes (around a 30% of maximum percentage over this region) and with those obtained by Gao et al. (2016) and Gröger et al. (2022), as other referees have pointed out. Thanks to the higher resolution of regional data, we were able to perform a local-scale analysis of the distribution of this AR-related precipitation percentage over the IP. It allowed us, e.g., to highlight a lower percentage over the Northwest due to a higher amount of non AR-related precipitation. In our study, instead of comparing a historical period and a future period, we analysed the changes in three simulations of the same period due to different aerosols treatments: prescribed (BASE), only direct and semi-direct effects included (ARI) and all aerosol-radiation-cloud interactions activated in the model (ARCI). Furthermore, as depicted in Fig. 5, the greatest percentage differences were observed in the ARCI-BASE comparison over the Southwest, showing an increasing of approximately 5%, which is not a exceptionally large difference.

Line 238: I am not convinced that 80 AR clusters is enough to overcome natural variability, could you add some discussion on the robustness of only using 80? Have you considered playing with your threshold to increase your sample size? Would the results be the same if you used a fixed-relative threshold, based on the “base” climatology? And/or a simple relative climatology unique to each of your experiments (base, ari, aric?) This would increase your sample size and also test uncertainty in your AR definition. (One thing that ARTMIP has shown is that the moisture threshold value is by far the biggest influence on AR frequency, and quite significantly so).

We kindly appreciate these interesting comments. We have played with the IVT threshold to see its influence on the number of ARs, their mean intensity and duration, the number of common AR events that would result and the percentage of AR steps shared by the three simulations. You can find a table displaying the results in the answer to your last question. Bearing that in mind, the 80 common events employed seem like a reasonable approach to extract conclusions, as we have clustered them based on their aerosol configurations. Increasing the sample size could have increased the number of members in every cluster, but the conclusions would have been similar.

The chosen threshold (300 kg m-1 s-1) is an absolute value (already discussed) that was derived from the computation of the 99th percentile (there was a typo in the manuscript that read "90th" instead of "99th", but it has been corrected) of the IVT over L1 in the BASE simulation, which yielded a time mean value of around 260 kg m-1 s-1. However, this percentile showed quite similar values (between 250 and 270 kg m-1 s-1) in ARI and ARCI, so the results could have been similar if relative thresholds were used.

Figure 6: I am not sure if this figure adds much to the manuscript as currently described. Their differences don’t seem significant by eye (?) How are they important? If they are not, then maybe omit this figure.

As we have answered to Referee #1, Fig. 6 shows the 80 common ARs events yet unclassified. More specifically, it shows the ARI-BASE (red) and ARCI-BASE (blue) differences in mean IVT, mean latitude and mean direction. As you have just pointed out, the differences seem like noise at a first glance (with the exception of the most intense AR events). Thus, the aim of this figure was to motivate and illustrate the need of the following EOF and clustering analysis to shed light on these differences, gathering similar events and then studying their relations with aerosols.

Figure 8: Add an explanation for the box and whisker styled plots: mean, median, quantiles? What is the color scheme showing? As clusters 2 and 3 are primarily discussed, perhaps only show these instead of all the clusters? It will be more focused.

You are absolutely right, an explanation of the box and whisker plots is missing and it may lead to some difficulties when interpreting the displayed results. For instance, one of the referees posed a question regarding what the red points (outlayers) were, because we had not mentioned them in the text. This explanation is going to be included in the revised manuscript.

The color scheme is just showing the ARI clusters/boxplots in different shades of red and the ARCI clusters/boxplots in different shades of blue, because red and blue colors represent these simulations along the work. It is just an aesthetic decision.

We have focused on clusters 2 and 3 because they were the ones that presented the biggest differences. However, we have discussed whether showing the rest of the clusters and we have concluded that it may be interesting to show how there is not a so clear signal in their differences.

Figure 11: Same comment as Figure 8, as well as only showing the significant clusters.

The response here is the same as the one to the previous comment, so we would like to refer to the answer there.

Thickness field diagnostic : Have you considered showing low level winds and/or IWV instead of the frontal boundaries via thickness field for these composite plots? I would think that IWV might be a better diagnostic to show ARs, given it is the moisture stream that makes the AR unique, and not all ARs are associated with the warm conveyor belt? To show strengthening/weakening of the thickness fields, the gradient value (i.e., anomalies ahead - behind the front might be more intuitive than the difference plots which are hard to interpret. I like the thickness plots for the case studies, which help to highlight the relationship between the AR and the strengthening/weakening of the frontal boundaries, but for the composites, they are hard to interpret. If difference plots are continued to be used, then significance should be added.

Thank you again for your comment. As previously explained, we have attempted to explore other variables but the uniqueness of the ARs obscured the underlying patterns. We would like to refer to the answer to the general comment 1 for a longer discussion about this matter and statistical significance, and to the answer to the "case studies" suggestion for the changes implemented in the thickness diagnostic plots.

Figures 8,11: There is a lot of information packed into these figures, but not alot of explanation in the text. Consider adding more description and inference with these figures to make your points.

Thank you for your comment. We will take it into consideration during the revision of the manuscript. Thanks to your previous suggestions and those of the other referees, the description and discussion of these two figures will be substantially improved and extended to make it as clear and complete as possible.

Figure 13: Contour labels need to be a bit bigger, it is hard to see them even after zooming in.

Thank you very much for the observation. We have made the contour labels bigger in Figures 13-18.

Case studies: I really like the figures with the dust and IVT overlays as this shows the displacements of the ARs. I would recommend trying to do something similar with the composites to help illustrate your conclusions that the aerosol locations and magnitudes impact intensity and location of the ARs. The case studies show this, but the current figures 8-12 aren’t as convincing.

Thank you very much for your comment. Following also the suggestion of Referee #3, we have added the trajectories of each AR belonging to a cluster and their mean trajectory, but instead of performing this approach to all the clusters and representing it on Fig. 7 and 10, we have focused on the most relevant clusters (discussed in the manuscript), and we have added the representation of the trajectories to Fig. 9 and 12, where the thickness fields are shown. For instance, you can find the resulting representation of the ARI clusters 2-3 in the figure below this paragraph. Each thin arrow represents an AR. It is located on its mean latitude with its mean direction and the length of the arrow is proportional to its mean intensity. The thicker arrow represents the mean characteristics of the ARs belonging to the cluster.

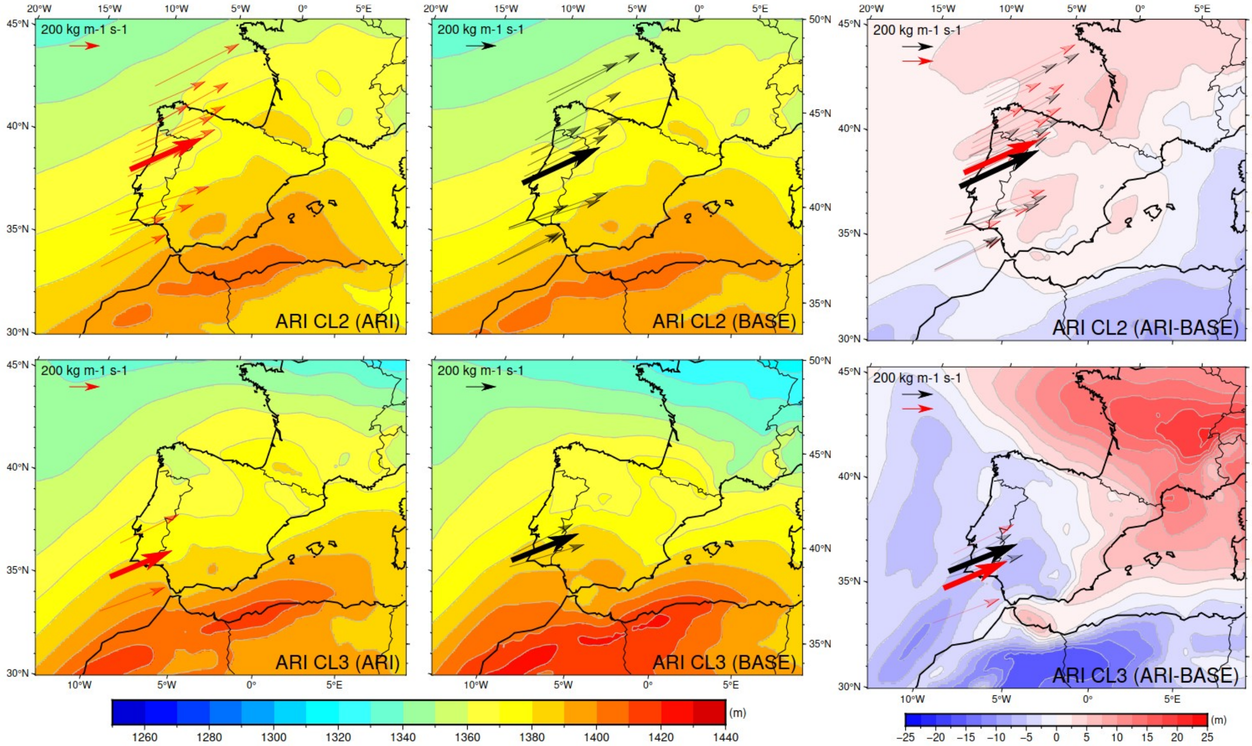


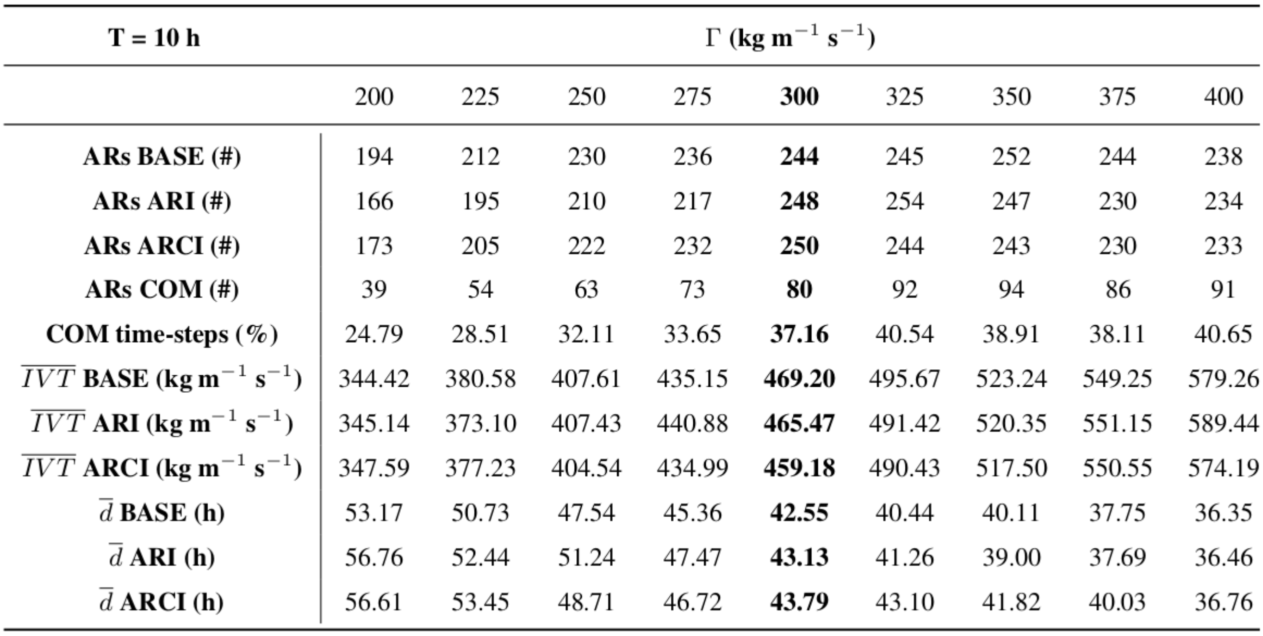
Figure: ARI and BASE mean thickness fields of the atmospheric layer between 1,000 and 850 hPa of the common AR events belonging to clusters 2 and 3 in the ARI simulation and ARI-BASE thickness differences. The same time steps are included in the representations of both experiments.

Line 342: This was not explained or motivated convincingly and AR uncertainty (that is, the uncertainty in AR metrics due to ARDT alone) is not addressed in the manuscript. This should be done given that AR frequency is highly sensitive to thresholding values. Suggested ways to address this: (1) Uncertainty can be discussed in the text addressing the limitations of using one ARDT, (2) For extra robustness and my recommendation, repeat the AR analysis by running the AIRA ARDT using different threshold values to both increase the sample size and attempt to bound ARDT uncertainty, (3) More work, but useful could be to compare AIRA with other ARTMIP ARDTs. Other ARDT catalogues for MERRA2 and ERA5 available, in addition to source data so AIRA could be run for a sample period for direct comparison. Data available at https://www.earthsystemgrid.org/dataset/ucar.cgd.artmip.html. Comparing to other regional ARDTs such as the IDL (Ramos), or the Brands ARDTs are highly recommended, especially if there are plans to use AIRA for other applications, including climate change where more than one ARDT is typically needed (O’Brien et al., 2022).

Thank you for all your comments. We have changed line 342 to specify that some of them are not suitable if the domain is so limited that spatial tracking can not be performed.

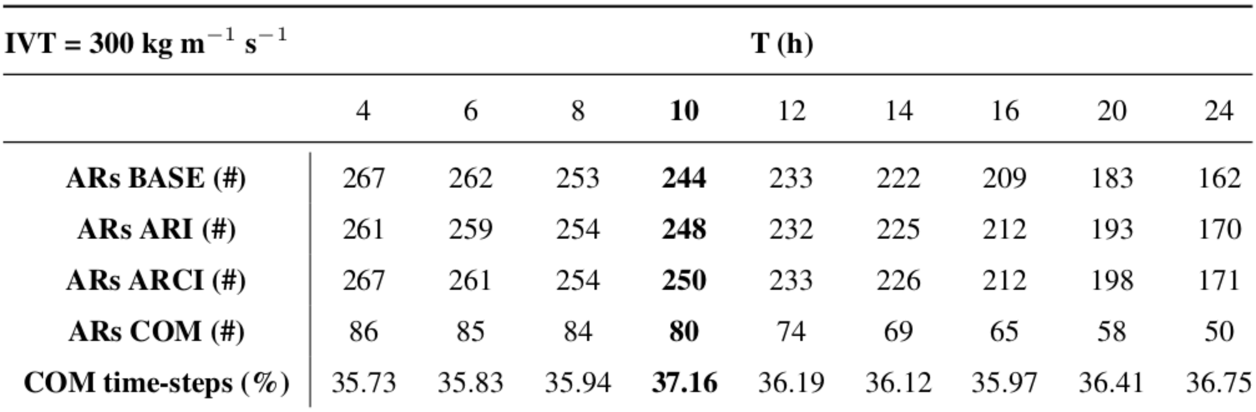
Referee #3 also mentioned the need of a discussion about the sensitivity to the threshold parameters. We have followed some of your suggestions. First, we have discussed in the text the limitations of using only one ARDT. Second, we have performed an analysis of the sensitivity to the IVT threshold given a fixed minimum duration and the sensitivity to the duration threshold given a fixed IVT threshold. The results are exposed below and include the variation in the number of ARs in each simulation, the number of common ARs events, the percentage of common AR time steps and the mean intensity and mean duration of the identified ARs.

Table: Sensitivity analysis to the IVT threshold, given a fixed minimum duration, of the number of ARs identified in the three simulations, the number of common AR events, the percentage of common AR time-steps and the mean intensity and duration of the ARs of the three simulations.



| **T = 10 h** | **$\Gamma$ (kg m$^{-1}$ s$^{-1}$)** | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | 200 | 225 | 250 | 275 | **300** | 325 | 350 | 375 | 400 |
| **ARs BASE (#)** | 194 | 212 | 230 | 236 | **244** | 245 | 252 | 244 | 238 |
| **ARs ARI (#)** | 166 | 195 | 210 | 217 | **248** | 254 | 247 | 230 | 234 |
| **ARs ARCI (#)** | 173 | 205 | 222 | 232 | **250** | 244 | 243 | 230 | 233 |
| **ARs COM (#)** | 39 | 54 | 63 | 73 | **80** | 92 | 94 | 86 | 91 |
| **COM time-steps (%)** | 24.79 | 28.51 | 32.11 | 33.65 | **37.16** | 40.54 | 38.91 | 38.11 | 40.65 |
| $\overline{IVT}$ **BASE (kg m$^{-1}$ s$^{-1}$)** | 344.42 | 380.58 | 407.61 | 435.15 | **469.20** | 495.67 | 523.24 | 549.25 | 579.26 |
| $\overline{IVT}$ **ARI (kg m$^{-1}$ s$^{-1}$)** | 345.14 | 373.10 | 407.43 | 440.88 | **465.47** | 491.42 | 520.35 | 551.15 | 589.44 |
| $\overline{IVT}$ **ARCI (kg m$^{-1}$ s$^{-1}$)** | 347.59 | 377.23 | 404.54 | 434.99 | **459.18** | 490.43 | 517.50 | 550.55 | 574.19 |
| $\overline{d}$ **BASE (h)** | 53.17 | 50.73 | 47.54 | 45.36 | **42.55** | 40.44 | 40.11 | 37.75 | 36.35 |
| $\overline{d}$ **ARI (h)** | 56.76 | 52.44 | 51.24 | 47.47 | **43.13** | 41.26 | 39.00 | 37.69 | 36.46 |
| $\overline{d}$ **ARCI (h)** | 56.61 | 53.45 | 48.71 | 46.72 | **43.79** | 43.10 | 41.82 | 40.03 | 36.76 |

Table: Sensitivity analysis to the minimum duration threshold, given a fixed IVT threshold, of the number of ARs identified in the three simulations, the number of common AR events and the percentage of common AR time-steps.



| **IVT = 300 kg m$^{-1}$ s$^{-1}$** | **T (h)** | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | 4 | 6 | 8 | **10** | 12 | 14 | 16 | 20 | 24 |
| **ARs BASE (#)** | 267 | 262 | 253 | **244** | 233 | 222 | 209 | 183 | 162 |
| **ARs ARI (#)** | 261 | 259 | 254 | **248** | 232 | 225 | 212 | 193 | 170 |
| **ARs ARCI (#)** | 267 | 261 | 254 | **250** | 233 | 226 | 212 | 198 | 171 |
| **ARs COM (#)** | 86 | 85 | 84 | **80** | 74 | 69 | 65 | 58 | 50 |
| **COM time-steps (%)** | 35.73 | 35.83 | 35.94 | **37.16** | 36.19 | 36.12 | 35.97 | 36.41 | 36.75 |

On one hand, a lower IVT threshold results in a decrease in the number of ARs but also in an increase of their duration, because two very close in time events could be identified as a single but longer event. On the other hand, increasing the IVT threshold over 300 kg m-1 s-1 reduces the mean duration of the ARs but has little impact on the number of ARs itself. For instance, the selection of an IVT threshold of 400 kg m-1 s-1 would have resulted in a decrease in the number of ARs in BASE, ARI and ARCI of 2.5%, 5.6% and 6.8%, respectively.

With respect to the sensitivity of the duration threshold, the results turned as expected. The higher the minimum duration imposed, the lower the number of ARs identified that meet this condition. Furthermore, we also wanted to remark that the selected parameter (T=10h), gives rise to the highest percentage of common AR time steps, with 80 common events that have allowed us to perform our comparison study.