Referee Comment #1

General comments:

This study is not the first that aim at detecting ARs in regional models. This should be mentioned. Some remarks are given in the special comments.

Thank you very much for your comment. 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 ARDT (Ramos et al., 2016) and Brands ARDT (Brands et al., 2017) 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 both 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.

The description of the algorithm should be improved and some deviation from existing ones should be explained.

As said right above, its deviation from the existing ones will be addressed in the revised manuscript.

Unlike others, the AIRA algorithm detects ARs on two longitudes at 10 and 12°E and infers additional length and direction criteria by employing trigonometric functions. Though this is described briefly in the text, a figure sketch with a zoom on L1 L2 to draw the trigonometric elements used to derive the relevant parameters would be helpful, e.g. the direction and length scales.

Thank you for the suggestion. A visual representation of the trigonometric elements used in AIRA will make the algorithm explanation easier to follow. We are going to include it as a new figure in the revised manuscript.

As far as I understand the IVT threshold was calculated using all time stamps and not only those at 12:00 UTC (when moisture is at the higher end) as in for e.g. Lavers et al. (2013). Likely this may result in a lower threshold which should be discussed.

The IVT threshold was derived from a calculation using all time steps. This concern is also present in the Special Comments section, so we would like to address the answer there.

The BASE, ARI and ARCI experiments should be better described for those readers who are not specialists in aerosol modelling and those who are not familiar with the WRF-Chem model. What precisely is meant by semi-direct and direct effects on a physical basis? The interaction of aerosoles with radiation beyond the optical depth in ARI should be physically explained. The same would help for the interaction of cloud (micro-)physics. Are condensation-nuclei reduced due to precipitation for example?). If so, in which of the BASE, ARI, and ARCI experiments is this the case? In the current version only references to literature about the WRF model and it's coupling is given. A brief summary about coupling prognostic variables, input, and output etc would be helpful. This knowledge is essential for the understanding of the results.

Thank you for your comment. These three experiments were developed and have been used before by other members of our research group. Their complete description can be found in

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 referenced this work in line 85. However, your comment made us realise that it may be not so clear that this reference intention was to offer the reader a complete description of the simulations. Therefore, we have change it to make it more explicit:

"Three experiments were considered in this study, each of which included different aerosol interactions. The complete description of these three simulations can be found in Jerez et al. (2021). "

Furthermore, we are going to include a brief description of the three experiments, answering all your question (aerosol-radiation interactions, cloud microphysics, CCN reduced due to precipitation), in the revised manuscript. In addition, a brief summary of the model inputs-outputs will also be included.

What is meant by direct effects (radiation scattering, absorption and emission) and semi-direct effects (thermodynamical changes in the clouds induced by direct effects) of the aerosols has been explained in the specific comments section.

In the results sections the physical processes that lead to differences in the three experiments should be better and more verbosely explained to meet a broader readership which are not only atmospheric or aerosole researchers. For example, often a heating or cooling is proposed but as no corresponding temperature anomaly is shown this is hard to see.

Thank you for your comment. Similar questions are found in the Specific Comments section, so we would like to address the answer there. However, the main explanation is that thickness is directly and solely related to temperature in an atmospheric layer between two fixed pressure levels. We have added a little explanation (see the answer to the specific comment) to the revised manuscript to make it easier to follow the results.

Also the clustering procedure which is based on leading EOFs of salt and aerosoles is not sufficiently described. All this makes it difficult follow the results and final conclusions. More examples are given below.

We have computed the sea salt and dust anomalies for a reduced spatial domain and then we have treated these two variables (the anomalies) as a single vector/field (of double the length of each aerosol field individually). We have then performed a Principal Component Analysis (PCA), also known as an EOFs analysis, of said field over time (considering the 80 common events). We have retained as much PCs (EOFs) as needed to explain at least a 75% of the total variance. Then, we have performed a hierarchical clustering (using the Ward method) over the PCA, and the centroids of the resulting clusters are shown in Fig. 7 and 10. Each cluster centroid consists on a dust field and a sea salt field.

This is also explained in the manuscript (lines 257-261): "Initially, an EOF analysis has been jointly performed for the sea salt and dust AOD (550 nm) standardised anomalies within the region bounded by -15°E and 4°E longitude and 33°N and 45°N latitude. The ARI and ARCI experiments used five and six EOFs respectively, explaining 75% of each total variance. A clustering classification was then performed on these analyses, which separated the common cases into eight different groups. The centroid of each cluster was associated with two centre fields, one per considered aerosol.".

Special Comments

line 7: “The analysis of common AR events showed that the differences between simulations

were minimal in the most intense cases, and a negative correlation was found between mean direction and mean latitude differences.

please rephrase: what is meant? you have three sensitivity simulations. When the ARs are located more to the North in e.g. BASE, then the direction is more south in ARI and ARCI? Perhaps it’s better to remove the second part of the sentence.

This sentence means that if an AR in ARI (ARCI) is located further North than in BASE (positive latitude differences), then the AR in ARI (ARCI) presents a more zonal direction (negative direction differences) and viceversa. An example sentence has been included in the abstract:

"This implies that more zonal ARs in ARI or ARCI with respect to BASE could also be linked to northward deviations."

line 11 deviations from what? What precisely is meant by reinforcement and attenuation? is it the moisture transport (in most studies taken as a proxy for intensity) or precipitation?

Deviations refer to spatial differences with the reference simulation, BASE. When we talk about intensity (or intensity reinforcement/attenuation), it refers to the magnitude (modulus) of the IVT. We've changed that line to make it as clear as possible:

"[...], inducing spatial deviations and IVT magnitude reinforcements/attenuations with respect to the BASE simulation depending on the aerosol configuration."

Introduction

line 33: what is meant by “anomalous”? Heavy precipitation above a certain threshold?

Yes, that's what is meant here. For more information, we recommend consulting the reference, where this statement was derived from.

Climate change is indeed assumed to impact on ARs. However, besides the important studies of Payne and Algarra, there also relevant studies with more focus on Europe and in particular the Iberian Peninsula. Please consider these to mention, like e.g.

Gröger, M., Dieterich, C., Dutheil, C., Meier, H. E. M., and Sein, D. V.: Atmospheric rivers in CMIP5 climate ensembles downscaled with a high-resolution regional climate model, Earth Syst. Dynam., 13, 613–631, https://doi.org/10.5194/esd-13-613-2022, 2022

Ramos, A. M., Tomé, R., Trigo, R. M., Liberato, M. L. R., and Pinto, J. G. (2016), Projected changes in atmospheric rivers affecting Europe in CMIP5 models, *Geophys. Res. Lett.*, 43, 9315–9323, doi:10.1002/2016GL070634.

Lavers, D. A., Allan, R. P., Villarini, G., Lloyd-Hughes, B., Brayshaw, D. J., and Wade, A. J.: Future changes in atmospheric rivers and their implications for winter flooding in Britain, Environ. Res. Lett., 8, 034010, https://doi.org/10.1088/1748-9326/8/3/034010, 2013.

Thank you very much for these references. We have found them very interesting and we are going to include them in the revised manuscript.

line 46: “tracking its long 2D structure”. Do you mean tracking its elongated 2D structure?

Yes, thank you for the comment. We have corrected it.

line 51: That’s true. The effect of resolved spatial orography on the representation of AR over land was found most evident over the Iberian Peninsula (see e.g. aforementioned study by Gröger et al.).

Thank you, we have added this citation here too.

line 57-61: The interactive online coupling between aerosole modules and other climate compartments will represent feedbacks by aerosoles in a much more realistic way. May be this could be explained a bit more in the Introduction. Can you mention some feedbacks we neglect if we use only prescribed fields of aerosoles instead of simulated ones?

As said in the General Comments, we are going to include a more complete description of the three simulations in the Data section. There, we are going to compare the BASE experiment (prescribed aerosols) with the ARI and ARCI experiments (aerosols calculation fully coupled in the model). However, in lines 57-61, we are trying to motivate/highlight the potential of these differences to change the simulated ARs. We find your comment very relevant, so we are going to extend these lines and mention some feedbacks that would be neglected in a prescribed aerosol configuration, like the changes in the CCN concentration due to precipitation or the modification of the cloud droplets properties based on the aerosol (acting as CCN) characteristics.

Methods

line 75: “The WRF-Chem model (v.3.6.1) was used for the simulations, both in a decoupled configuration (WRF alone (Skamarock et al., 2008)) and in a fully coupled configuration with atmospheric chemistry and pollutant transport to account for aerosol-radiation and aerosol-cloud interactions (Grell et al., 2005)"

What does fully coupled mean and how is the coupling precisely done in the three experiments? This is essential to understand the results in this study. The section could benefit from a brief description of the WRF-Chem and how aerosoles have direct and semi direct effects on the models physics.

Thank you for your comment. This brief description of the WRF-Chem and aerosols effects will be included in the revised manuscript as addressed on a previous answer. Fully coupled means that it is included as an active coupled component; i.e., the model chemistry (aerosols) is computed simultaneously and integrated into the dynamics of the WRF model. This contrasts with the stand-alone configuration, in which the results of both parts can be computed independently and the chemical part is not re-introduced into the model.

section 2.1 Data

line 81: "... encompass major dust emission areas". Which are these areas? The Sahara desert?

Yes, the Sahara desert and its surroundings are the main dust emission areas for the IP. An explicit mention has been added to the manuscript.

line 87: what is CCN and how does it interact with model physics?

CCN stands for *Cloud Condensation Nuclei.* Their concentration and nature alter the physical properties and amount of cloud droplets, thus changing the lifetime of the clouds and the thermodynamics of the atmospheric layer in which they are present. Further information about the microphysics scheme used in the simulations will be included in the revised manuscript, in the brief experiments description mentioned before.

aerosol-radiation interactions: does radiation then alter the optical properties of Aerosoles and or the number of condensation nuclei?

Radiation generally does not alter the optical properties of the aerosols, but aerosols do alter radiation by means of scattering, absorption and emission processes (that is what is called *direct effects* of the aerosols). These processes can also induce thermodynamical changes in the clouds (*semi-direct* effects), altering the size of the droplets and/or their development.

line 131: "First, the magnitude and direction of the IVT are bi-linearly interpolated to the detection lines, L1 and L2, enabling the computation of the required variables. "

What variables are meant here? The sentence implies that IVT is calculated from specific moisture, u, and v as a first step and thereafter IVT is interpolated from the models grid to L1 and L2. What variables do you mean here in addition to u,v, and q and for what are they necessary?

Thank you for your question. IVT is calculated only from q, u and v. When we said "required variables", we were refering to the derived variables that can be computed from the IVT vector and the geometry of the detection lines, such as the direction, the width or the IVT maximum in both lines at a given time step. These variables are required later in the algorithm to determine whether an AR has been identified.

We have changed the sentence to make it more clear:

"First, the IVT magnitude and direction are bi-linearly interpolated to the detection lines, L1 and L2, enabling the computation of the geometrical and physical variables required later in the algorithm."

Line 134: How is the threshold value determined? Is this threshold latitude dependent? Is it determined from climatological values like e.g. the 85th percentile as in other algorithms? L1 extents over a wide range of latitudes ranging from semi-arid climates to more wet conditions. Are the northern latitudes more represented in the threshold than those from the south?

The threshold value is an absolute value stablished by the user. It is not latitude dependent and it is not determined by computing percentiles, at least in the algorithm itself. However, we recommend computing them beforehand to decide the threshold. For instance, we have chosen an IVT threshold of 300 kg m-1 s-1, based on the 99th percentile value of the IVT on L1 (260 kg m-1 s-1). As for the L1 question, the detection line 1 extents over a wide range of latitudes but we do not think that any of them are more represented than the others. In fact, this methodology is also applied by other ARDTs. In the figure below, we show the distribution of the mean impact latitude of the identified ARCI ARs (similar results were found for the other experiments), which turned to be more or less even.

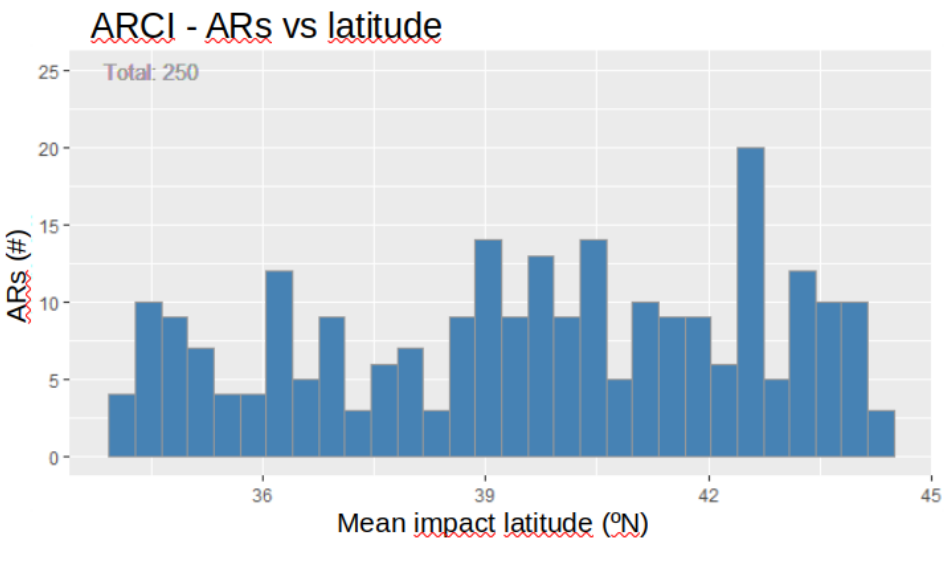

Figure: Number of ARs versus their mean impact latitude in the ARCI simulation for the whole period.

Line 139. “...direction of AR...”. If I interpret equation 4 right, wouldn’t the term orientation not better than the term direction? Direction might be related more to the movement of the AR over time.

Thank you for the interesting remark. From our perspective, both terms (direction and orientation) would be correct in this case, because the ARs tend to move in the direction given by their orientation.

Line 165: How is “s” determined? Do ARs not move over time so that changes in their axis latitudinal position are not unusual? Please explain why this is necessary.

The parameter "s" is determined by the extention of the detection lines and prior knowledge of the AR behaviour in this area, where the occurrence of two consecutive ARs is not so rare. ARs move over time, thus changing the latitudinal position of their maximums (spine), but these changes are gradual due to the movement. If we detect a large shift in the position (almost the length of the detection line), we assume that an AR is passing by the South and another one is arriving to the North, or viceversa, of the dectection area, instead of consider both of them as part of the same AR event. Distinguishing these events as different ARs is the reason why this parameter is needed here and adopts such high value. However, it can be changed by the user.

line 169: “.. estimation of the AR length..”. Do you mean AR duration here? The length scale isn’t determined so far, is it?

The consecutive time steps mentioned before in that sentence correspond to the duration of the AR (or at least an estimation of it), but this duration allows the estimation of the AR length, knowing the wind velocity. That is what this sentence refers to. In this paragraph, our aim was to just present the algorithm, thus no specific correspondence between length and duration has been made yet for the studied area. In section 3.1 (AIRA implementation to our study domain), one can read the following: "Secondly, the minimum time duration for an interval to be classified as an AR was set at T = 10 h. Given that the mean wind speed associated with ARs in the study area is around 30 m s^{-1}, this minimum duration would indicate the occurrence of an AR of approximately 1,000 km in length."

185 ff:

Please explain how the value of the mean 90^{th} percentile is calculated. Is it determined over all latitudinal points (i.e. m=22) at L1 and over the whole time period? Or do you calculate 22 90 percentiles an at the end average over the 22 points? Also contrary to other algorithms you take into account all day times while others include only time steps of 12:00 UTC time stamps (when moisture content is high due to solar heating). This is likely the cause why your value of 260 kg $m^{-1}s^{-}$ seems a bit lower than in other studies (see e.g. Lavers and Villarini, 2013: https://agupubs.onlinelibrary.wiley.com/doi/full/10.1002/grl.50636)

Thank you for your question. First of all, we would like to clarify that there was a typo in this sentence, and we have already amended it in the revised manuscript. The 99th percentile was what was supposed to be written here. This percentile was just used to give us an idea of the IVT threshold and it was computed as the mean 99th percentile of the IVT field over L1 during all the considered time steps, i.e., for every given time, we computed the field percentile, and then we applied a time mean. With respect to your second remark, we used all the time steps, instead of only those with the 12:00 UTC time stamp. As you pointed out, this is likely why the 260 kg m-1 s-1 seems lower than the threshold obtained by percentile calculations in other studies. However, we were aware of this issue and thus we chose a higher IVT threshold, 300 kg m-1 s-1. In the revised manuscript, we will provide a brief discussion about this matter, referencing other studies.

Is there any empirical evidence to support the limits for w (150 – 800 km)?

There are different ARs catalogues, and some of them even include a representation tool to see the identified ARs. At the beginning of our research, we explored these catalogues and came up with these limit values for the AR width, which seem reasonable. A minimum width of 150 km allows us to distinguish very thin structures that are not ARs. However, we want to clarify that these limits are just parameters of the algorithm and thus are adjustable by the user.

The spatial / temporal criteria listed in Table 1 seem to be more or less reasonable from theoretical/geometrical considerations, but ultimately lack empirical evidence. So it would supportive if sensitivity tests could be made to estimate the sensitivity of the thresholds on the AR frequency, duration and intensity. If this is too much effort, this should be at least discussed in terms of uncertainties associated with the algorithm.

Thank you for your comment. This remark was also mentioned by the other referees. Following your suggestions and those of the other two referees, we have performed an analysis of the sensitivity of the IVT threshold given a fixed minimum duration and the sensitivity of the duration threshold given a fixed IVT threshold. The results are exposed in the tables 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	Γ (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
$\bar{d}$ **BASE (h)**	53.17	50.73	47.54	45.36	**42.55**	40.44	40.11	37.75	36.35
$\bar{d}$ **ARI (h)**	56.76	52.44	51.24	47.47	**43.13**	41.26	39.00	37.69	36.46
$\bar{d}$ **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.

section 3.2.1

Figure 5 fits very well with result from Gao et al. (Fig. 8) and Gröger et al., 2022 (Fig. 5d). They could be mentioned to support the validity of the new developed AIRA algorithm.

Gao et al.: https://journals.ametsoc.org/view/journals/clim/29/18/jcli-d-16-0088.1.xml

Thank you very much for the comment. We have added the references of both studies to the revised manuscript.

section 3.3 Common events

you may consider renaming the section, e.g. coherence of events or so

We kindly appreciate your comment. This is something that we have already discussed and we ended up choosing "Common AR events". We would like to keep the section name as it is, because the main idea is that we have analysed here the AR events shared by the three simulations, thus we think that the term "common", common to the three simulations, fits well in this case.

I would speculate that the different treatment of aerosoles will alter not only the precipitation pattern of AR related precipitation events but also alter systematically the mean precipitation rates. Could it be that the alteration seen in AR related P are similar to those in mean P? Implying that aerosoles impact similar mean and AR precipitation events.

Thank you very much, this is a very interesting question. Of course, the different aerosol treatments may affect not only the AR-related precipitation but also the non AR-related precipitation distributions. And they affect it indeed. However, if aerosols had affected both precipitation distributions exactly similarly, there wouldn't have been changes from simulation to simulation in the percentage of the total precipitation that could be related with ARs, as the alterations would have been similar and thus compensated in the computation of the percentage.

3.3.1. Analysis of differences

What is the idea of eliminating non coherent AR intervals to elaborate the effect of aerosoles? I think here a more profound explanation for the strategy should be added. From a methodological point of view I would guess ARs penetrate into the EuroCordex model domain roughly at the same time and at the same position. Then, differences in precipitation, IVT intensity and frequency etc. would be attributed to the different treatment of aerosoles. Can you confirm this? Consequently, the non coherent AR time steps would be the result of the aerosole treatment which would neglected in this approach. Would it be wrong to calculate Fig. 6 without the eliminating step?

Thank you very much for this interesting insight. The idea of eliminating non coherent AR intervals was to do a one-to-one comparison between the ARs of the simulations. Therefore, we needed to have the same time steps to compute the differences and then compare the IVT intensity, mean trajectory, etc. For instance, you wouldn't be able to calculate Fig. 6 without the elimination step, because the number of identified ARs is not the same in all the simulations, thus e.g. AR #150 may not be the same AR in ARI and ARCI. This is why we have followed this methodology. However, there are other approaches, like the one you mentioned. It would consist on identifying each AR in the three simulations and comparing for instance their "arrival time". We computed something similar a while ago, during the first stages of our research. Not only did it make the analysis way more complex but also it led to some dead-ends, due to the impossibility of relating the results with the effects of aerosols, at least at the general scale pursued in our research. We concluded that this approach could be very interesting in a single case study but it fell out of the scope of this work, where general conclusions have been found and study cases were used as illustrative examples of these conclusions.

line 245: “...The maximum IVT was obtained by averaging the maximum IVT magnitude of each AR event in the three experiments…”. You mean intensity here?.

Thank you for the comment. Yes, we meant the maximum intensity of each AR, which corresponds to the maximum IVT magnitude of the AR, as explained in the second paragraph of section 3.2. However, we have modified this sentence to make it more clear:

"The mean maximum IVT was obtained by averaging the maximum IVT intensity of each AR event in the three experiments."

line 246: What is meant by spatial deviation. Does it refer to the deviations in latitude (Fig. 6 middle)? Please be consistent with the terms throughout the manuscript.

When we refer to 'spatial deviations,' we are addressing differences in latitude and direction; i.e., we are referring to all non intensity related differences. We have added this aclaration to the manuscript.

line 247: what is meant with “the three magnitudes”. A distinction between three magnitude categories was not done before.

Thank you for your observation. We meant "the three variables" shown in the figure. We have modified the sentence, which now reads as follows:

"The spatial deviations (latitude and direction differences) tended to zero, and the ARI-ARCI differences of the three considered variables (latitude, direction and mean IVT) became minimal in the most intense events."

line 249: Can you summarize Fig. 6 to explain what you aim to analyze with the EOF analysis. Are there systematic differences in the deviations to BASE in Fig. 6? At a first glance, it seems like noise (with the exception that most intense ARs seems to be consistent in the experiments). Also, it would be interesting to show at least the first or three leading EOFs for sea salt to get an impression where most variance is concentrated.

Thank you for your question. 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 commented, the differences seem like noise at a first glance (with the exception of the most intense AR events). Thus, the aim of the following EOF and clustering analysis was to shed light on these differences, gathering similar events and then studying their relations with aerosols. The EOF analysis (or PCA) was primarily used to reduce the dimensionality of the study problem and to perform the clustering (hierarchical classification, see the answer to the related question in the General Comments section for more information) over it, to obtain the different groups of events which were subsequently studied individually.

With respect to the first or three leading EOFs, we also find it interesting and we are going to add it to the revised manuscript as an appendix, or maybe as supplementary material.

section 3.3.2

Please explain more verbose how the EOF analysis was performed, e.g. how were salt and aerosole anomalies calculated, what clustering algorithm was applied etc. Moreover, is there a seasonality in the aerosol fields (as you showed for AR incidents)? Could the clustering also explained by over-representations of certain seasons? Did the classification procedure require to choose the number of different clusters? If so, why were 8 classes chosen?

As previously mentioned in the General Comments section, we have computed the sea salt and dust standardised anomalies for a reduced spatial domain, treating these two variables as a single vector/field (of double the length of each aerosol field individually). We have then performed a Principal Component Analysis (PCA), also known as an EOFs analysis, of said field over time (considering the 80 common events). In fact, the PCA function in R is able to perform also the computation of the standardised anomalies, making the analysis straightforward. We have retained as much PCs (EOFs) as needed to explain at least a 75% of the total variance. Then, we have performed a hierarchical clustering (using the Ward method, which is the default method in the HCPC function in R) over the PCA. This classification procedure has an optimal number of resulting clusters (obtained by elbow diagrams), but one can choose a different number of clusters by looking at the tree diagram of the classification. That's how we made our decision (see Figure below). The centroids of the resulting clusters are shown in Fig. 7 and 10. Each cluster centroid consists on a dust field and a sea salt field.

With respect to the aerosols seasonality, we want to clarify that we have included in the analysis only the aerosol fields of the 80 common events, and these events are not evenly distributed along the year. Autumn common AR events are the most numerous. In all the 8 clusters of each experiment (ARI and ARCI), we have found AR events of very different seasons. For example, ARI cluster 2 gathers 1 January AR, 3 March ARs, 2 April ARs, 1 from May, 1 from June and 4 October ARs.

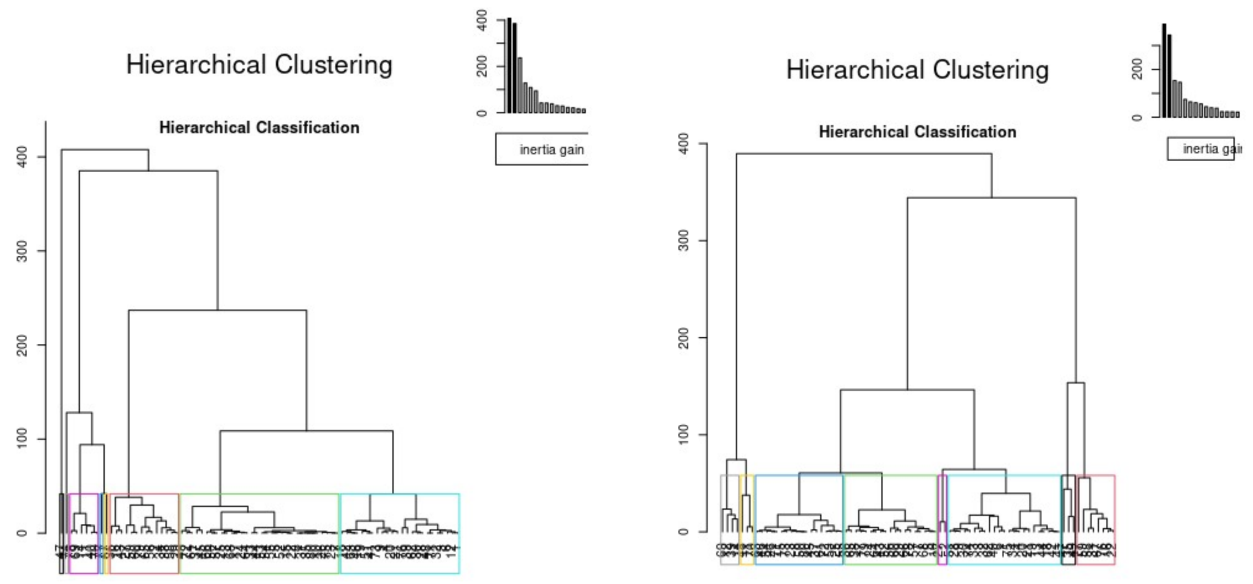

Figure: Tree diagrams of the hierarchical clustering classification of sea salt and dust aerosols jointly in the ARI (left) and ARCI (right) experiments.

line 259/Figure 7: “ ...it was observed that an AR weakening occurs in clusters 2 and 3.” Not clear at first reading what is meant. Figure 7 top (which I think this statement is related to) shows the IVT difference at the y-axis, in the caption it reads as “magnitude”, and in the text it is termed weakening.

The magnitude of a vector is also called the modulus of a vector. IVT magnitude refers to the modulus of the IVT, the intensity of the AR. The mentioned statement is related to Figure 8 top, where the y-axis shows the IVT differences and the x-axis shows the different clusters. As you can see, the IVT differences between ARI and BASE in clusters 2 and 3 were mainly negative, thus the IVT of the ARs related to the ARI simulation was generally lower than the IVT of the ARs related to BASE. Therefore, the ARs in ARI are weakened in comparison with BASE.

Figure 7 shows red points which are not explained.

Is it possible that you are referring to Figure 8? Figure 8 is a boxplot and the points represent the outlayers. They are more present in the clusters with a higher number of members due to a higher variability between the members of those clusters. However, Referee #2 suggested the inclusion of a little explanation about what a boxplot shows (quartiles, mean, outlayers, etc.), so there will be a more profound insight in the revised manuscript.

line 262: what kind of frontal surface? that of a storm? An explanation at this first place what is meant by a thickness to non-specialists is lacking.

We strongly appreciate your observation. An explanation of the thickness of an atmospheric layer and its relation with temperature is clearly missing and it may cause a missconception of the present section for those readers who are not so familiar with this definition. Before further discussion of the thickness field and answering your first question, ARs are usually related to the frontal surfaces (cold fronts) of extratropical cyclones, as said previously in the Introduction section.

In meteorology, the thickness of an atmospheric layer refers to the vertical distance [m] between two pressure levels, which define the layer. The hypsometric equation (see equation below) represents the relation between the thickness (x) of an atmospheric layer and its mean temperature (T). R_d is the specific gas constant for dry air, g_0 is the standard gravitational acceleration, P_1 is the pressure of the inferior level and P_2 is the pressure of the superior level.

$$x = \frac{R_d T}{g_0} \ln\left(\frac{P_1}{P_2}\right)$$

Therefore, the thickness of the layer is directly and solely related (by a multiplicative constant) to its mean temperature given two fixed pressure levels (1000 and 850 hPa in this study). That is why we are able to talk about cooling or heating by analysing the thickness fields. The higher the thickness, the higher its mean temperature.

We have added an explanatory sentence and the citation to a well-known book of meteorology, so that non-specialists can follow the results presented in this manuscript while having access to a further and more detailed reading about this topic:

"ARs are commonly associated with a frontal surface, which can be identified by analyzing the thickness field. The thickness field of an atmospheric layer is directly and solely related to its mean temperature given two fixed pressure levels, as depicted in the hypsometric equation (Stull, 2011)."

line 267: “In cluster 3, a wider cooling effect is present, but the more pronounced cooling in the south (over the north of Africa) leads to the observed weakening”.

Where is this cooling derived from? Figure 9 show thickness [m]. There is no information about temperature differences at this place.

Thank you again for your question. It is derived from the thickness differences given two fixed pressure levels. As explained before, in such cases, the thickness of the atmospheric layer is directly related to its mean temperature.

Accordingly I have to go back to Figure 7 where an elevated dust concentration in the region is visible. Shall I interpret this as proxy for cooling in this region (in the sense of dimming?). So far no explanation for the assumed cooling is given at the place of line 267. I get lost here...

Thank you for your comment. We are sorry for the confusion. Although there is not a complex explanation for the cooling, we should have made the statement in a more explicit way (and we have corrected it in the revised manuscript). The cooling is mainly due to the scattering and absorption of solar radiation, also known as direct effects or aerosol-radiation interactions. In a not so summarised way, the explanation would be as follows. In the BASE experiment, the AOD is set to zero, which means that radiation encounters a perfectly "clean" atmosphere. In contrast, the ARI experiment includes the on-line calculation of the aerosol optical properties (AOD) and their interactions with radiation are activated in the model. With that said, in the ARI experiment, part of the incident radiation is scattered and absorbed by dust aerosols, thus changing the mean temperature of the considered atmospheric layer (1000-850 hPa) with respect to the BASE experiment. These dust-related temperature changes are usually a warming of the atmospheric layer in which the aerosols are present and a cooling of the surface and its adjacent atmospheric layer.

line 279 ff: The ARCI-BASE comparison reads much better than the previous ARI-BASE comparison because physical explanations that appear plausible are given to the reader. This should be likewise provided for the ARI-BASE comparison. Saying this, most of the explanation is based on the interpreted aerosole effect on temperature which is not shown itself, though often it is argued with “cooling” or “warming”. Therefore it would help to show additional plots for temperature either instead of thickness or as supplementary material.

Thank you very much for the positive feedback about the ARCI-BASE comparison. Thanks to your previous comment, and as said in its answer, we are going to add an explicit explanation for the cooling found in the ARI-BASE comparison to make it more complete. With respect to the temperature plots, we refer again to the definition of thickness and its relation with the mean temperature of an atmospheric layer. Bearing this in mind, we do not find necessary to show temperature plots, because all our considerations can be derived and followed by means of the thickness plots. Furthermore, they include information of a whole layer instead of only representing a specific pressure level. However, if you still find it necessary to include some additional temperature plots (perhaps for two or three different pressure levels) after all the previous considerations, we could include them as supplementary material.

3.3.3 Case studies

line 309: “...-70.32 and 58.01 kg m−1 s−1...” over which area has this been averaged. Over the AR area? Model domain? Iberian Peninsula?

Thank you for your question. It is not a spatial average, but a time average. Each AR is characterised by its mean intensity (mean IVT modulus of the AR spine), among other variables. Thus, when we mention the IVT differences between the simulations, we are just comparing the mean IVT of that AR in the three experiments.

lines 309 to 333:

This paragraph reads very well as it provides a process-based discussion about the aerosol effects on ARs, involving a chain of interactions between temperature, clouds, droplets etc. The role of heating/cooling and temperature gradients is again highlighted and the reader may wonder if it would be possible to support this statement by a figure showing e.g. temperature anomalies.

Thank your for your feedback about this paragraph. Once again, we want to refer to the hypsometric equation and how it directly relates the mean temperature of an atmospheric layer with its thickness. For additional information, see the answers above.

In particular, the paragraphs (and already the previous ones) emphasize the cooling effect by aerosoles as well as a heating effect from more abundant droplets, prolonged cloud presence, and latent heat gain are discussed. However, it becomes not quite clear why the individual effects (cooling or heating) dominate in the respective cases. This could be more explained.

Thank you very much for your comment. We have shown the effects of each aerosol type (dust and sea salt) when only their interactions with radiation were implemented and when both aerosol-radiation and aerosol-cloud interactions were activated in the model. We have seen that microphysics effects tend to dominate over radiation effects in the ARCI simulation, not only compensating but also surpassing the radiative effects. This may be due to their relation with greater energy (heat and thus temperature) changes.

line 323-324. Isn’ it rather a southwestward shift seen in Figure 15 ARCI-BASE?

Thank you for your remark. That sentence was only referring to the mean impact latitude of that AR over the detection line. That is why we just said "southward" shift/deviation, because the line has a fixed longitude. We have changed this sentence to make it more clear in the revised manuscript.

Conclusions

Including an atmospheric chemistry and trajectory model yields likely the most realistic and physically consistent treatment of aersoles. But it is likely also the most expensive? If so can we derive from the experiments a statement whether or not the additional online coupling of an expensive chemistry/aerosole model is worth and/or in which cases? Can we expect systematic shifts in AR related precipitation and or moisture convergence which may be of importance on climate related time scales? Would the conclusions also hold for e.g. the U.K. which is further away from major dust aerosole sources?

Thank you for these interesting questions. Including the atmospheric chemistry and aerosol transport in the model is 4 to 8 times more expensive, according to the experts in our research group. On one hand, from a physical point of view, the more accurate representation of the physical processes leads to more realistic interactions between the model components and thus better and more realistic results. The higher computational cost can be worth in researches that aim to study these physical processes and/or the relative significance of each interaction. This was the case of the present work. On the other hand, some studies suggest that the differences obtained in the most expensive and most physically realistic runs are mainly relevant at the very local scale (distribution changes). Therefore, these very complex simulations would not represent a substantially better reference for operational use at the synoptic scale.

With regards to the shifts in AR-related precipitation, our findings reveal that they would be very case-dependent, influenced by the aerosol fields configuration present at that moment. In this research, the distinction between dust and sea salt aerosols effects was made. Therefore, the conclusions regarding sea salt aerosols would also hold for the U.K., even if the dust aerosol concentration is negligible.