Summary

In this study, the authors investigated the future changes of compound explosive cyclones (ECs) and atmospheric rivers (ARs) in the North Atlantic using the simulation data from six CMIP6 models. Different from previous papers, this study focused on the future changes of compound ECs and ARs, which usually develop rapidly due to the strong diabatic feedback and are closely related to extreme precipitation and wind. The authors found that there is a significant and systematic future increase in the EC-AR concurrences, especially over Western Europe in the high-emission scenario. Overall, this study investigated the future changes of ECs and ARs from a novel perspective and the paper is well written and organized. However, I have some major concerns, especially for the methods, and general comments listed below for the authors' consideration.

First of all, we would like to thank the anonymous reviewer for the helpful comments on the manuscript.

Major Concerns:

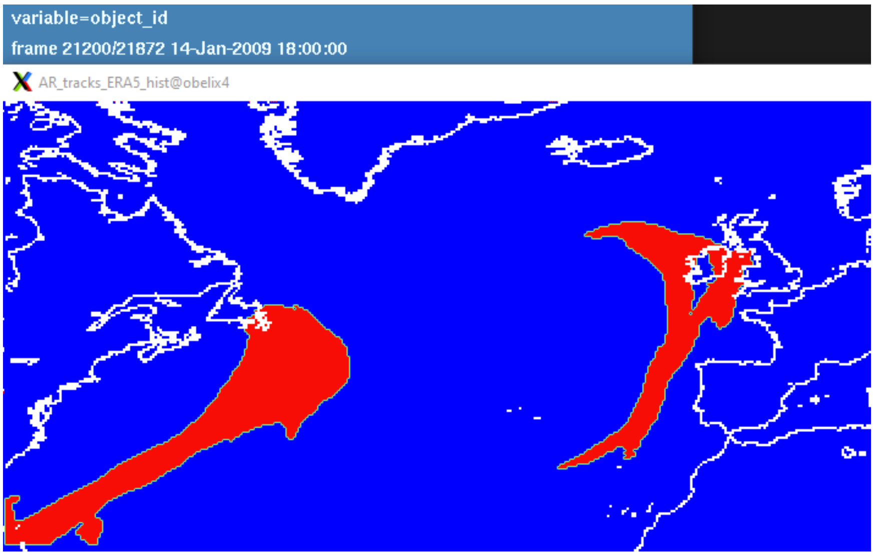
(1) My biggest concern is the sensitivity of the conclusions in this study to the AR and cyclone tracking methods. For example, there are many different AR detection methods with large differences as summarized in some papers from the Atmospheric River Tracking Method Intercomparison Project (ARTMIP; e.g., O'Brien et al. 2020, 2022; Shields et al. 2019). If a different AR detection method or a different cyclone tracking method is used, will that have a significant impact on the conclusions about the EC-AR concurrences?

Thanks for this comment. We are aware and acknowledge that tracking algorithms can show significant differences (see the last paragraph in the conclusions). Among other reasons detailed in the manuscript, we choose to use the TempestExtremes algorithm to detect and track both cyclones and ARs because the cyclone and AR climatologies shown in Figures 1 and 2 sit in the middle of the results for the climatologies of cyclones and ARs shown in the tracking comparison studies of Neu et al. (2013) (for cyclones) Collow et al. (2022) (for ARs). Despite that, we tested the sensitivity of our results by comparing those with Eiras-Barca et al. (2018), our methodologies only differ by the use of different tracking algorithms. They use two different tracking algorithms and ERA-Interim, and our results for ERA5 of the rate of concurrence (Figure 4 a,b) sit in between their results for the two algorithms (Eiras-Barca et al. (2018), Figure 3a). We are aware that the absolute values of concurrence rates can be sensitive as shown when compared to Eiras-Barca et al. (2018). Still, the changes between historical and future scenarios need to be tested with the same methodology. The main objective of our paper is to asses future changes in the rate of concurrence of cyclones and ARs and for that same methodology should be applied to both present and future periods. Given this focus, we believe that differences among periods are less sensitive to the tracking algorithms. This is supported by Zhang et al. (2024) in Figure 7 where they show larger agreement in the AR frequency future changes than in the climatologies among different AR tracking methods. A similar result is shown by O'Brien et al. 2022 (Figure 1), where differences in

AR climatologies among tracking algorithms are large but those agree in future AR frequency trends.

(2) Section 3.3: "an extratropical cyclone (EC or non-EC) is linked to an AR by detecting the presence of an AR within a 1500 km of the cyclone center" (lines 133-134). It may be oversimplified to use 1500 km distance to determine the concurrence between cyclone and AR. If a cyclone and an AR are dynamically associated with each other, the AR is usually located over the south to southeast side (the position of the low-level jet stream ahead of the cold front) of the cyclone center. In that case, it makes sense to define the concurrence if an AR exists within 1500 km. However, if an AR is located over the north or northwest side of a cyclone center, I don't think it is reasonable to say that the AR is dynamically associated with the cyclone even if the distance is within 1500 km.

Thank you once again for your insightful comment. Regarding the choice of the 1500 km threshold, I kindly refer you to our response to the first comment in our reply to Mika Rantanen. In addition to the sensitivity tests already mentioned there, we will also test the quadrant in which the closest AR grid point detected around the cyclone lies. Pending this analysis, we underscore that there are examples where the closest AR grid point to a cyclone can sometimes be located to the north or northeast of the cyclone and still have a dynamical association with it. The shape of the AR, particularly its front part, often takes on a sickle shape that curls around the cyclone. As a result, it may not be uncommon for the closest AR grid point to be in these quadrants and have a dynamical association with the cyclone. Next, we provide an example of the AR "sickle shape" (from ERA, 14th January 2009 at 18 UTC):



(3) Fig.4, the description in Section 4, and many other places throughout the manuscript: It is a smart way to use the maximum deepening point (MDP) as a reference point. However, the lifetime of extratropical cyclones has a large variability, varying from a couple of days to over one week. So it is very arbitrary to say that 36 hours before MDP (-36 h) is "the initial stages of cyclone formation", +36 h is "the dissipation stages of the cyclones", and from -36 h to +36 h is "the lifetime of the ECs".

Thanks for this comment, other reviewers also pointed this out. We agree that formation/dissipation stages might not be appropriate as we analyse the cyclones +/-36 hours from the MDP. We will change this and modify the text accordingly to avoid referring to formation/dissipation stages in this context.

(4) The domain for analysis is 25N-65N and 80W-10E in this study (line 74). Did the authors use only the data within this domain for AR/cyclone tracking and EC-AR concurrence determination? If yes, there will be a boundary issue, especially for the cyclones and ARs near the boundaries, since both AR and cyclone detections have thresholds for moving distance and existing time. For example, in Fig.1 the EC track density is unreasonably high along the western boundary of the domain; in Fig.8 the rate of coincidence tends to be very small close to the boundaries. My concern is that the boundary issue may have impacts on the conclusions, especially for Western Europe, the area around the south tip of Greenland, and other places close to the boundary.

Thank you for this important comment. This point has also been raised by another reviewer; please refer to our response to the second comment from Reviewer #2.

(5) For the horizontal resolution, ERA5 is 0.25 degrees while the six GCMs are quite different, varying from ~0.7 degrees to ~2.0 degrees. Did the authors interpolate the data to a common grid before analysis? Will the different horizontal resolutions have any impact on the conclusion? For example, the AR intensity is defined as the maximum IVT, but the different horizontal resolutions may have an impact on the maximum IVT across different models and ERA5. As a result, some differences across different models and ERA5 might be a data resolution issue, not the real model bias.

We did not interpolate the data to a common grid; all analyses were conducted at the native resolution of the datasets. This approach is consistent with how Ramos et al. (2016) handled different grid resolutions in CMIP5 when analysing ARs or similar to O'Brien et al. (2022) and Zhang et al. (2024) that apply the AR tracking algorithms to the CMIP models native grids. We acknowledge that differences in model resolution introduce uncertainty, but we consider this an inherent uncertainty of the models themselves. Moreover, when analysing differences between periods, we compare the multi-model means, which is equivalent to comparing each model to itself and then averaging the differences. We never perform a direct comparison across different models. Our main goal here is to assess changes between historical and future scenarios of AR intensity and not model evaluation with ERA5. We acknowledge that ERA5 shows the largest differences with the historical models, so we will study the sensitivity of resolution with ERA5 and compare it to the historical CMIP6 models.

Finally, we recognize that maximum IVT is particularly sensitive to resolution. However, we chose this variable because it is not dependent on the AR shape. An alternative would be to use the mean IVT within the AR, but AR shape (particularly the size) is also sensitive to model resolution.

(6) A few concerns/questions about the AR and cyclone tracking methods.

Line 118: In addition to detecting ridges in the IVT field, there is an IVT minimum threshold of 250 kg/m/s. However, the IVT values have large variability from low to high latitudes. Will the 250 kg/m/s minimum threshold be too high for the ARs at high latitudes, like the area around the south tip of Greenland, near or higher than 60N?

The 250 kg/m/s minimum threshold was added as a sanity check as it is one of the most used thresholds, but the detection threshold used is the laplacian of the IVT, less sensitive to latitude differences. Anyway, 250kg/m/s is not a high threshold for ARs at least under 65N (our case), see ARs IVT climatologies in Thandlam et al. (2022).

Line 118: The AR candidates should have an area larger than 4 × 10^5 km^−2. Are there any requirements for the AR shape? ARs are usually defined as a long and narrow corridor of strong water vapor transport.

In this case, the geometrical requirement is the area threshold (also a very common threshold in AR detection algorithms). The TempestExtremes algorithm does not have a wide/length requirement, even the many of the AR tracking algorithms have a length requirement this is this is not the case for all (Shields et al. 2018).

Line 119: "... concatenated if at least one grid point …" Similar to major comment (5), if the data were not interpolated to a common grid, it is a concern since models have quite different horizontal resolution, which means the "one grid" threshold is different across different models (~0.7 degrees to ~2.0 degrees) and ERA5 (0.25 degrees).

We acknowledge that this can be a source of uncertainty, but in our answer to comment (5) we consider the model resolution as an inherited source of uncertainty of the model itself.

Lines 91 and 92: "not exceed 6 GCD degrees" and "at least 12 GCD degrees". Are there any specific reasons for using 6 GCD and 12 GCD?

For these parameters, we used the default tracking setting from Ullrich et al. (2021).

Minor Comments

(1) Line 11: "worst-case scenario", I think it would be better to use high-emission scenario.

Thank you for your comment, we will consider it.

(2) Why did the authors select those six CMIP6 models while there are many other GCMs in CMIP6?

We kindly refer you to our answer to the 7th comment of Referee #2.

(3) Line 101: "These results agree with Priestley and Catto (2022) and Zappa et al. (2013) …" It's worth noting that Zappa et al. 2013 used CMIP5, not CMIP6.

We will mention it in the text.

(4) Line 103: “Figure 1c,d”, do you mean Figure 1 c and f?

Yes, we will correct that.

(5) Line 134: delete “a” before “1500 km”. Thank you, we will correct the typo.

(6) Fig.1 and Fig.2: It would be very helpful to show the percentage difference in addition to the absolute difference of cyclone track density and AR frequency. Same for the other difference figures.

We will consider changing the absolute difference for the percentage difference in these figures.

(7) Fig.2: The unit of AR frequency is % in the figure but the values are fraction (0.00~0.12).

You are right. We will fix it.

(8) Fig.4: The difference across different lines (models) is not very clear. Maybe use different colors for different models?

Thanks for the advice, we will try to find a set of colors that make the different lines more distinguishable.

(9) Line 144 and some other places: “MDP point”. “Point” is redundant since MDP is maximum deepening point.

You are right, it is redundant. We will remove “point” after the MDP in all places.

(10) Line 149: “… favours the detection of an AR in its surroundings”. “detection” is not suitable here, maybe change it to “existence”.

We agree, will change the word.

(11) Line 158: “the standard deviation of the rate of coincidence”. Is that calculated using the coincidence rate at MDP or from -36 h to +36 h?

It is calculated using the coincidence rate at each time from -36 h to +36 h.

(12) Line 173 and 175: 0.08 and 0.05 are the model biases of what? Coincidence rate? For the average of all models or the model with the maximum bias?

Yes, of the coincidence rate. It refers to the model with the maximum bias. We will rewrite these sentences to make them easier to understand.

(13) Line 227: “The AR intensity for ERA5 is larger than any model for the historical period because ERA5’s resolution is almost 4 times higher than the CIMP6 models, and attains larger values of IVT-max.” So the difference of AR intensity between ERA5 and models might be a data resolution issue? This is the same question as my major concern (5).

Thanks for your comment, we agree and acknowledge that one of the main reasons for the difference in the IVT-max in this case (ERA5 vs CMIP6) might be driven by the resolution difference. We would kindly refer to our answer in the comment (5).

(14) Fig.6: Does the Non-EC AR means the ARs without an explosive cyclone or the ARs without any cyclone (no matter weak or explosive)?

It means ARs with a non-explosive cyclone (non-EC). The ARs analysed here are those associated with ECs or non-EC.

(15) Fig.8e, there are many areas show a large increase in coincidence rate with model agreement. Why did the authors only emphasize the western Europe in the Abstract (line 12)?

We emphasize the Western Europe area in the abstract because it is potentially more relevant for the possible impacts in the society as being an area highly populated.

(16) Fig.8 a and b, the brown-green color map gives the readers the impression that the brown and green areas are opposite rather than low to high. I would suggest using a different color map.

Thanks for the suggestion, we will consider using another color palette.

References:

Neu et al. (2013): https://journals.ametsoc.org/view/journals/bams/94/4/bams-d-11-00154.1.xml

Collow et al. (2022): https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2021JD036155

Zhang et al. (2024): https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2023JD039359

Thandlam et al. (2022): https://link.springer.com/article/10.1007/s00704-021-03776-w

Shields et al. (2018): https://gmd.copernicus.org/articles/11/2455/2018/

Ramos et al. (2016): https://agupubs.onlinelibrary.wiley.com/doi/full/10.1002/2016GL070634

Ullrich et al. (2021): https://gmd.copernicus.org/articles/14/5023/2021/