We thank the three reviewers for reading our manuscript and for providing us with interesting comments that can lead to an improved version of the manuscript. We provide a point-by-point response to all the comments below. Reviewers' comments appear in blue and our response in black.

Reviewer 1

The study uses the recently developed CANARI-LE large ensemble to analyse four storms that affected the UK and to compare their characteristics under the SSP3-7.0 (Regional Rivalry). They found that the total number of cyclones similar to such storms decreases in the future and the most extreme events intensify in terms of precipitation and low-level wind speed.

The manuscript builds upon the cyclone analogue methodology introduced by Ginesta et al. (2024), and this prior study is acknowledged in the main text. However, beyond applying the framework to a different model and a new set of UK-impacting storms, it is not yet clear what additional conceptual or mechanistic understanding is gained. If the primary contribution is to assess robustness across models and storm types using this methodology, this should be stated explicitly — already in the abstract and reiterated in the conclusions — for example, by framing the study as a systematic extension aimed at testing generalisability and moving toward broader or more automated application of this analogue framework.

We agree that the objective of the paper should be much clearer. Indeed, the purpose of our study is not to repeat the work by Ginsesta et al. (2024). Instead, our aim is to generate new insights about the possible changes in impactful storms under climate change and to point towards improvements to the methodology. We will revise our manuscript to ensure the paper's objectives are clear and consistent throughout.

A clearer comparison of the results with Ginesta et al. (2024) and, most importantly, its implications, would strengthen the manuscript. Although the storms are different and not directly comparable, the authors should discuss more explicitly how their results relate to the CESM-based findings: do you find similarities with specific storms, are magnitudes or scaling different, are there model-dependent sensitivities, are the results so storm-dependent that no generalisable conclusions can yet be drawn? This would help clarify the added value of the study and if it extends or generalises the previous work meaningfully.

We agree that this is a very good suggestion. A clearer discussion of the relationship between our results and those by Ginesta et al. (2024) will be included in a revised version of our manuscript.

The ensemble size (1200 model years per period) should adequately sample internal variability. However, since the analysis relies on selected analogue events, it would be useful to check whether the analogue selection favours particular modes of variability (e.g. ENSO, NAO, AMOC). Even a check for one storm would increase confidence in the robustness of the results.

Some of these modes of variability are available for the LE and it will therefore be relatively straightforward to determine the distribution of the analogue storms with respect to those modes of variability. We will consider whether to include this analysis in the revised version of the manuscript. However, it is likely that particular phases of the modes of variability are favoured by the analogue identification. For example, the NAO's positive phase is associated with increased storm activity over the UK and Ireland and therefore it is possible that this phase is better sampled by the analogues.

The number of analogues appears small, and for some storms very small. Why do the authors use a constant 500 km radius along the entire track, rather than a varying radius (e.g. narrower near peak intensity and wider when the storm is farther from its peak)? The authors state that sensitivity tests were performed without finding significant differences; however, if using a 700 km radius increases the number of analogues and thus the statistical sample, why not adopt that choice? In other words, how do the authors define the point at which a storm ceases to qualify as an analogue?

Whether a storm qualifies as an analogue or not depends on the arbitrary threshold chosen. The purpose of introducing a fixed threshold for analogue identification is to avoid selecting storms that are not close to the target storm. In the methodology's original version by Ginesta et al. the threshold was different between storms and between periods (historical v future). Applying it to our set of cases had the effect of selecting storms that could not be considered analogues of the target storms and their quality could differ between the present and the future. However, we acknowledge the reviewer's point. Therefore, we will revisit our choice of threshold to either increase the radius or further justify the value used, including results from a sensitivity analysis. Note that our method requires (1) track maximum intensity location within 300 km of the location of that of the target track, and (2) mean track separation at the preceding four 6-hourly time steps (i.e. 24 hours) is less than 500 km. In this sense, while the threshold is fixed, analogue tracks are free to deviate from the target track within these restrictions.

Precipitation: why do the authors not consider the total accumulated precipitation over the entire storm track? Alternatively, analysing 6-hourly precipitation along the track (e.g. as a histogram, cumulative distribution, or time series relative to peak intensity) could provide additional insight and account for potential shifts in the timing of max precipitation. Do precipitation changes follow Clausius–Clapeyron scaling? Is the intensification mainly thermodynamic, dynamic, or a combination of both?

Considering total precipitation along the storm track and considering 6-hourly as a time series are both good suggestions. We will seek to implement them in the revised version.

Please clarify whether analogues are restricted to winter (or extended winter, or September to April as stated in a few figures). If so, this should be clearly stated; if not, possible seasonality effects should be discussed.

The analogues are not restricted to winter. Some of the storms show clear seasonality but others don't. We will consider including seasonality effects in the revised manuscript.

Definition of extreme analogues: why do you use the 90th percentile based on ERA5 and not a specific percentile according to the model, as Priestley and Catto 2021 did? This threshold might not be the same in CANARI LE.

The purpose of this choice is to quantify changes in frequency of cyclones that would be classified as extremes under present climate conditions.

Other comments:

Please use colour-blind friendly colorbars (avoid rainbow)

Use blue for precipitation increases and red for decreases.

Adjust colorbars to correspond only to the plotted fields (e.g., Fig. 2)

Regarding the three comments on colourbars, we will re-consider the choice of colourbars and colourbar range in all the figures following the Reviewer's recommendations.

Please clarify what is meant by "dynamical intensity" in the third research objective.

We use the phrase 'dynamical intensity' to refer to a measure of cyclone intensity in terms of purely dynamical characteristics, such as relative vorticity, and distinguish it from an intensity measure based on fields, such as wind speed or precipitation, whose effects pose a direct hazard to society.

Reviewer 2

The authors build on an existing approach to study changes in archetype cyclones through studying changes in their analogues. A considerable part of the paper is dedicated to evaluating the CANARI LE dataset. The paper is generally well-written. The analysis is logically structured and provides detail on the analysed cyclones.

What I am more concerned with is the interest that the paper may hold for the broader WCD readership. Given that the paper slightly modifies an existing approach and applies it to a different dataset and different cyclones, I would encourage the authors to clarify the value of the paper for those not involved in CANARI or not interested in the specific geographical subregion being studied. This is partially done in the conclusions, but more could be done in this respect in the introduction and discussion sections. For example, the authors could reflect on the feasibility of systematically applying this approach to a very large number of cyclones (would this be computationally feasible? Would the results be interpretable?) or on whether the approach could work for other atmospheric dynamical features beyond cyclones.

We thank the reviewer for their valuable suggestions for further discussion. We will incorporate them in a revised manuscript. Indeed, we want the manuscript to convey the generality of our findings beyond the underlying dataset or the region of interest.

I provide some additional specific comments below.

1) Title: Given that, according to Fig. 1, at least two of the four cyclones being analysed appear to cause stronger winds over France than over the British Isles, and a third mainly affected the Republic of Ireland, the authors may consider revising their title to include both the British Isles and France, rather than only mentioning the UK.

The title will be revised to reflect this comment.

2) l. 9: This sentence makes it sound as though the authors are drawing a general conclusion on all cyclones. Perhaps they could state explicitly that this and the following points apply to cyclones similar to the four events that they are analysing.

The sentence will be rewritten to make this clear.

3) ll. 148 and following: It is worth mentioning here that reanalyses often heavily underestimate the local magnitudes of extreme events, including heavy precipitation and strong winds.

This will be mentioned in a revised version of the manuscript. We will add references to support these comments, e.g. Lavers et al. (2022, https://doi.org/10.1002/qj.4351), who point out that while ERA5 is able to represent location and distribution of heavy precipitation events, it is not able to represent the actual precipitation totals, or Gandoin and Garza (2024, https://doi.org/10.5194/wes-9-1727-2024), who discuss the underestimation of offshore wind speeds.

4) Sect. 3.2.2: You state that you identify analogues across 1200 simulated years of present and future climates, from 1980 to 2010 in the present-day climate period, and from 2070 to 2100 in the future climate period. However, in Sect. 5 you also identify analogues directly in ERA5. I would ask the authors to verify whether this apparent discrepancy issues from my own misunderstanding of the information or is indeed a case of incomplete information being provided in Sect. 3.2.2.

It is the latter. We will make it clear that in addition to finding analogues in the climate model simulations we also find them in ERA5.

5) Consider whether Figs. 4 and 6 could be combined into a two-panel figure.

We will consider combining these two figures.

6) Consider adding vertical labels to each row in Fig. 7, to indicate which cyclone is being shown without having to refer to the caption.

Thank you for the suggestion. Following it will greatly improve the readability of the figure. We will add labels in a revised version of this figure.

7) Sect. 6: In general, I see a very weak link to hazards throughout this section. In the typical view of risk, hazards should be quantities that may be directly related to impacts. Here, the authors seem more intent on analysing quantities of dynamical interest than quantities directly associated with impacts. The only part of the section that truly deals with a hazard is the subsection looking at precipitation. I provide some concrete examples of this below.

Precipitation and wind at 850-hPa, as an estimator of the maximum surface wind gusts are both impactful quantities, but we agree that the Section's title is currently misleading. We will reorganise the discussion so that dynamical and hazardous fields are better separated.

Sect. 6: Most studies of cyclone-related hazards focus on the hazards over land, where the overwhelmingly largest exposure is located. Here in some cases the largest changes in the quantities that the authors look at are over the ocean, making these quantities of limited relevance from a risk perspective.

There are applications for which ocean conditions are the most relevant (shipping, fishing). For other applications, conditions over land are crucial. However, our approach is to consider that the modelled cyclones could have occurred on slightly different locations. Thus, rather than considering the modelled surface effects, we consider the potential hazard that the target storms' analogues represent. This is also the motivation to investigate 850-hPa winds as these are less affected by contrasts in surface roughness between land and ocean.

l. 323 What is the logic of using maximum RV to study hazards? Would it not be more informative to select the day of largest cumulative hazard over land regions (quantified through whichever metric the authors may prefer)?

Maximum RV is not really being used to study hazards. Instead it is used to study the dynamical intensity of the storms, i.e. as a measure of cyclone intensity in terms of purely dynamical characteristics as opposed to an intensity measure based on fields, such as wind speed or precipitation, whose effects pose a direct hazard to society. This will be clarified in a revised version of the manuscript. As mentioned above, we will also reorganise the discussion in this section to make sure that the distinction between dynamical and hazardous fields are clearly separated.

Sect. 6.2: I would not call wind at 850 hPa a hazard. If the authors want to investigate hazards, they should focus on 10m wind gusts, 10m winds if gust data is not available from the model, or similar.

850-hPa wind speeds are used here as an estimator for the maximum wind surface gust. This will be clarified and supported in a revised manuscript.

8) Statistical significance testing; In some of the figures, the authors conduct statistical tests at individual gridpoints and assign a significance level to these. As this is a repetition of the same test a large number of times, I would ask the authors to confirm that they have applied a multiple testing correction to prevent spurious statistical significance results. If this has not been applied, then it is necessary in an eventual revision of the paper.

In performing statistical significance tests we were following the methodology of previous literature (e.g. Ginesta et al. 2024). However, we will consider the Reviewer's suggestion when revising the manuscript.

9) l. 388 I find "exceptionally unique" an odd turn of phrase. Unique is not a comparable adjective.

'Exceptional' is probably sufficient in this case.

10) Dataset availability: Could the authors specify whether any researcher can obtain access to JASMIN or whether this is restricted to e.g. UK-based researchers or researchers participating in specific projects?

The CANARI LE dataset is now available through the Centre for Environmental Data Analysis (CEDA) which is open to any researcher. The manuscript will be updated to reflect this.

Reviewer 3

Review of "Climate change effects on analogues of contrasting extratropical cyclones over the UK"

The manuscript by Morgan et al., investigates the influence of climate change on four historical cyclones impacting the UK, using cyclone track analogue approach. The study utilises the CANARI Large Ensemble, assessing both the present climate (1980–2010) and a high-emission future scenario (SSP3–7.0, 2070–2100). By focusing on these four cyclone case-studies, the study provides insights on cyclone-specific responses, which may be important for assessing their regional impact. Results show that track analogues exhibit contrasting responses to anthropogenic warming and explores the drivers of changes in intensity and meteorological hazards.

Overall, the manuscript is well written and presents some interesting results on projected changes in extratropical cyclones under anthropogenic warming. There are however some points that need to be addressed. My suggestion is for minor revisions, as detailed in the following.

Major comments:

This study is situated within the CANARI (Climate Change in the Arctic-North Atlantic Region and Impacts on the UK) research programme (Schiemann et al., 2026). As such, a significant part of the manuscript (e.g., analysis shown in figs. 3,4,5,6) is dedicated to a general evaluation of model biases and model response to climate change. The rest is dedicated to a process-based analysis of four selected cyclone analogues.

It is however not clear to me how the analysis of four cyclone analogues should be interpreted in the context of model biases in the CANARI LE simulations. This dual purpose of the paper affects structure of the paper and perhaps masks the key findings. Particularly, in several places throughout the manuscript, it states that the goal is to verify the performance of the CANARI LE in simulating North Atlantic cyclones, while in other parts, including the abstract, the research goal is to examine the influence of climate change on four contrasting historical cyclones impacting the UK.

I would suggest better justifying why this study is needed and how adopting a UK-impact perspective allows us to gain new insights, which were not in the original study. Re-organizing the sections concerning the model biases can be better integrated (e.g., plots 3-6 (or some of them) can move to the supplementary, to help the paper become more concise and focused on its process-based analysis of these cyclones in present and future climates.

We understand the Reviewer's concerns. The purpose of including relevant discussion about the performance of the CANARI LE at simulating North Atlantic cyclones was to provide a theoretical background to inform the results from the analogue analysis. We will address this concern in a revised version of the manuscript by making clearer our primary objective, which indeed is examining the influence of climate change on four contrasting cyclones over the UK. We will rewrite the relevant sections to make clear that the analysis of the performance of the CANARI LE is ancillary to the primary objective. We will consider whether any figures can be moved to the supplementary material.

Novel aspects of this work: In what aspect does this study go beyond the findings of Ginesta et al. (2024), given that the same methodology and approach are used (replacing CESM1 with CANARI LE)? The manuscript can strongly benefit from a comparison of the results with exciting CESM (which was used in the Ginesta et al., 2024 study) and from clarify how using the LE improved the analysis. This can be added to the discussion, providing a clearer context to this work and strengthening the manuscript. Specifically, it would be interesting to know if model biases in cyclone frequency, intensity, extremes storms (if possible, cyclone analogues) show similar results. A deeper discussion or analysis can help to generalise the conclusions beyond the 4 cyclone analogues and enhance the contribution of this study.

This point was also raised by Reviewer 1. A clearer discussion of the relationship between our results and those by Ginesta et al. (2024) will be included in a revised version of our manuscript.

Sample size of analogues: Across the 74-year period of ERA5, relatively few analogues are found for each of the selected cyclones (14, 27, 17, and 5). This makes evaluating statistical significance a challenge for the cyclone track analogue approach. Is there a way to increase the number of analogues? Would it affect the results? Would a selection of analogues based on percentile (e.g., top 10% of matches) can increase the number of selected cyclones in these cases? A sensitivity test or discussion about the Methodology can help to clarify this issue.

The low numbers of analogues in ERA5 reflect the limitations incurred when analysing a single realisation of the current climate. However, it is not desirable to artificially increase the number of analogues as this could lead to the addition of storms that don't share the characteristics of the target storms. Instead, we address this limitation through the analysis of the CANARI Large Ensemble. We investigated the use of percentiles in the identification of analogues. However, we found that following this procedure also allowed the identification of storms with little resemblance to the target storms. Therefore, we decided to use a fixed distance threshold.

In addition, the definition of extreme analogues is based on the maximum RV higher than the 90th percentile of cyclones in the Northern Hemisphere between 1980 and 2010 in ERA5. This threshold can be adapted perhaps for the North Atlantic or Euro-Atlantic sector, and can be implemented using model climatology rather than ERA5.

This point was also raised by Reviewer 1. Indeed the threshold can be adapted in several ways but it will always be arbitrary. The purpose of this choice is to quantify changes in frequency of cyclones that would be classified as extremes under present (or near present) climate conditions.

Minor/technical comments:

Abstract: "such projections average together cyclones with a range of contrasting dynamical characteristics potentially obscuring climate change effects on particular types of cyclones and the airstream structures within them." This sentence remains a bit unclear to me. What makes the cyclone characteristic contrasting?

By contrasting we mean that they follow an unusual path (e.g. Arwen) or gave a particular origin (e.g. Ophelia). We will clarify this in a revised version of our manuscript.

Abstract: Are the airstream structures discussed in the paper? Perhaps rephrase.

Airstream structure is not part of our analysis. The phrase has been removed.

Fig.2 shows candidate-track density for the Great Storm of 1987. Not sure if this figure is needed, and perhaps can be moved to the appendix?

We will redesign these and other figures showing differences. We will also consider the Reviewer's suggestion about moving this to the supplementary material section.

Figures 4 and 6 can be combined into one (multi-panel) figure, to allow a better comparison on historical vs. ERA5, and future vs. historical.

This was also suggested by Reviewer 2. The figures will be combined into a single figure in a revised manuscript.

Scientific phrasing: I suggest rephrasing the text in a more scientific writing style. Some examples: "Ophelia is included because..", "likely exacerbated damages", etc.

We will carefully revise the manuscript to ensure that the language is as scientific as possible.

Line 110: "This trajectory likely exacerbated damages to infrastructure.." Is it possible to examine the actual range of the loss? (for example, in Met Office reports or other sources such as Perils/EM-DAT should have some loss estimations and discussion on sources of damage).

Reference:
https://www.perils.org/files/News/2017/Loss-Annoucements/Ophelia/PERILS-Press-Release-Ex-Hurricane-Ophelia-27-Nov-2017.pdf

Thank you for the reference. We will quote the loss estimation from the Perils report.

Line 82: south of Englance -> England?

Corrected.