Authors' Response to Reviewer 2

General Comments. This paper aims to produce cyclones classes which capture various aspects of cyclone intensity better than metrics that use a single diagnostic. Overall, the paper describes the motivation and methods used well. There are however several parts of the paper that require clarification, particularly the clustering method, and more work is required to illustrate how future studies can implement the recommendations made. These points are described below in more detail and should be addressed before the paper is suitable for publication in NHESS.

Response: We would like to thank you for the feedback and the valuable comments which helped improve the quality of our manuscript. We have carefully addressed all the issues item by item as follows.

Comment 1

My main concern is that the 4 clusters identified by the authors are not straightforward to implement by others in future studies. The authors recommend that 5 variables are considered, but how to combine these to identify different classes of cyclone in future studies is not clear. Will subsequent studies need to repeat the Gaussian mixture model with their own tracked cyclone data to identify the 4 clusters? Also, the clusters are identified from the sPCA figure, so that step would also need to be repeated I believe. I would like the authors to provide a more step-by-step guide to how the intensity measures should be combined to 'comprehensively and non-redundantly quantify the intensity of ETCs' (lines 645 and 659). If the cyclone clusters are to be used to see how different kinds of ETCs respond to climate change (line 688) simpler instructions are needed on how to create them.

On a related point, on line 692 the authors state that their method 'allows a vast amount of information to be condensed to a level that is manageable for operational forecasters', how would the forecasters use the information? Is it envisaged that they would be provided with the 'intensity' of a cyclone based on a score from each of the clusters, or do the authors have something else in mind?

Response:

To address the issues about reproducibility and applicability of our analysis to other datasets, we have added a short example script to the Zenodo repository with the trained sPCA model included. A reader can use this model and their own data (or our data) to see to where any storm or object which can be described with the 11 intensity measures falls in the sPCA space (Cornér et al., 2024). A script has been provided which does the same for the Gaussian mixture model (GMM) as well. This script predicts the clusters of objects which can be described with the reduced set of intensity measures, i.e., VO, WS850, PRECIP, WFP, and SSI. Mentions of these models have been added to the text as well. We propose that this set of five intensity measures is "comprehensive and

non-redundant", i.e. it describes the intensity from all relevant aspects and does not contain the same information twice. These measures cannot be combined in the literal sense of the word, but should be used together when quantifying ETC intensity. For example, we claim that investigating only vorticity, one cannot draw conclusions about the increase or decrease of ETC intensity.

The clusters are not identified from the sPCA figure. The clusters are produced with the GMM which is run with the reduced set of intensity measures. This reduced set is identified by using the sPCA result and correlations between the intensity measures. This has been clarified in the text.

To answer the second part of the comment: Our framework could be used in operational forecasting to assess the uncertainty of possible storm impacts. Although we reduce the set of intensity measures to five measures, this is a lot of information to investigate when one uses an ensemble prediction system with possibly tens of ensemble members. We suggest that the information could be condensed to a manageable level by first determining the cluster of a storm in each ensemble member by using the trained GMM instance and then seeing how much disagreement there is in the ensemble. This would offer information on both the intensity and possible impacts of a storm as well as how uncertain this estimate is. This of course requires additional information such as predictions of storm locations and is dependent on the selection of points of interest.

Specific comments:

Comment 1

Line 11: What do the authors mean by 'impactful storms'?

Response:

By impactful storms we refer to extratropical cyclones which had a societal impact due

to e.g. heavy precipitation-associated flooding or other damage to infrastructure from winds.

Comment 2

Line 35: Vorticity is a noisy field typically including both mesoscale and synoptic scale features. Are the authors referring to a filtered vorticity field when they say that vorticity metrics describe the synoptic scale dynamics of ETCs?

Response:

This does in fact refer to filtered vorticity which represents synoptic-scale features. The filtering is explained later in the text in Section 2.2. We have decided to keep the text in the introduction unchanged as we do not want to introduce technical details in it.

Comment 3

Line 46 and 54: What is meant by 'concise' metrics? If the aim is to produce a concise metric, this should be defined.

Response:

We have revised the sentence in the text to clarify what is meant by concise metrics as follows:

> Secondly, a manageable number of metrics which are easy to compute (concise metrics) are needed to identify whether any trends in ETCs intensity have already occurred or may do in the future as the climate changes.

We also want to highlight that producing a single concise metric is not an aim of the paper and therefore avoid using the singular form of the word in this context.

Comment 4

Line88: What is T1639?

Response:

T_L639 refers to the spectral truncation of the grid in model which has been used to produce ERA5 and thus the output resolution of ERA5. It is a linear triangular truncation in which the largest total wavenumber that can be represented in the grid is 639. See explanation e.g. here: https://confluence.ecmwf.int/display/OIFS/4.3+OpenIFS%3A+Horizontal+Resolution+and+Configurations.

Comment 5

Line 102: What is the consequence of ignoring the biases in ERA5 data? Are you results sensitive to these biases?

Response:

When using a reanalysis, it is impossible to avoid biases in the data. All reanalysis datasets have some biases and choosing between reanalyses means choosing between biases. We determined that ERA5 is the best suited for our study in the North Atlantic–European region as it has been shown to perform well in many aspects related to ETCs and is easily available.

It is difficult to estimate the consequence of ignoring biases or the sensitivity of the results to these. We can speculate that the underestimation of high precipitation values in ERA5 may cause our precipitation distribution to be too narrow. This may have an effect on the cluster analysis through e.g. creating more overlap in precipitation between the clusters. We have added a mention of this is the discussion. It would perhaps be beneficial to perform similar analysis with a different reanalysis dataset but it is outside the scope of this paper.

Comment 6

Line 125: What are the time steps referred to here?

Response:

The time steps refer to the time steps in the data. Two days equals 16 time steps with 3-hourly data. This has been clarified in the text.

Comment 7

Line 151: '... available as is in the reanalysis.' This sentence does not make sense to me.

Response:

The comment has been answered also in Comment 12 of Reviewer 1 which states: The sentence has been reformulated from

> and are available as is in the reanalysis.

to

> and can be obtained from the reanalysis with no or minimal post-processing.

Comment 8

Line 163 and 177: What is the consequence of a mismatch or even no match between the location of the vorticity maxima and mslp minima? Is the latter a consequence of the fact that vorticity can capture the early stages of cyclone development before a closed isobar is identified in mslp?

Response:

The consequence of a mismatch between the location of vorticity and MSLP extrema is that the wind field is investigated around the vorticity maximum whereas traditionally the closed MSLP minimum might be considered as the reference for an ETC centre. However, in most cases the distance between these two extrema is smaller than the distance between the vorticity maximum and the wind speed maxima, which means the associated wind speed values are largely unaffected. We do not believe that the latter is a consequence of the lack of a closed isobar. The MSLP minimum does not need to be closed but just needs to be a local minimum.

A sentence has been added to clarify how the tracks with no found associated MSLP value are dealt with:

> For a small number (3 %) of maximum VO values, TRACK is unable to find an associated MSLPa value. These ETCs are omitted from the dataset.

Comment 9

Line177, 330: Here and elsewhere the authors refer to ERA5 wind gusts. It would be useful to have a brief explanation of this diagnostic quantity and how it is derived? Why is it underestimated in some areas (line 206)?

Response:

Wind gusts in ERA5 are calculated as the sum of the 10 m wind speed, a term accounting for surface roughness, and a term representing the contribution of convective downdrafts (Equation 3.99 in ECMWF, 2016). Wind gusts are underestimated in some areas because one or more of these terms is underestimated. The underestimation occurs mostly in regions of complex orography and/or areas with high wind speeds (Chen et al., 2024; Minola et al., 2020). This is briefly discussed in the Discussion section of the paper.

The details of ERA5 wind gust and how we use it were originally included in the manuscript but were removed for the sake of shortening the text. A reference of how wind gust is calculated in ERA5 is included in the paper (Bechtold and Bidlot, 2009).

Comment 10

Table 1: What is the difference between accumulated and time-integrated?

Response:

This is briefly explained at the end of section 2.3.2. The precipitation values are pre-processed by summing together 1-hourly values to obtain an accumulated precipitation rate per 3 hours. Therefore, we can just sum the values together to get an accumulation throughout the whole ETC track. The SSI values are, however, represented only every 3 hours as instantaneous values and cannot be summed together in the same way as precipitation values that are accumulations. Instead of calculating the accumulation by summing together the instantaneous values every three hours, we integrate them with respect to time, i.e. multiply the difference between respective time steps with a time interval of 3 hours. This ensures that the accumulated SSI value is (nearly) independent of the temporal frequency the data are available at. We use the word "accumulated" for both SSI and precipitation since the interpretation of both measures is similar: they quantify the relevancy for impact across the whole track.

Comment 11

Line 252: Do the authors use the 'type' of correlation, i.e. linear or non-linear later in the analysis or interpretation of their results? I may have missed this. Does the fact that the MI correlations are higher for SSI than the Pearson correlation imply that they are non-linearly related to the windspeed for example?

Response:

The fact that the correlation between the dynamical intensity measures and SSI is non-linear, is used to justify the inclusion of SSI in the reduced set of comprehensive intensity measures. The reasoning behind this is that as opposed to the wind-related measures, which are strongly correlated with each other, the non-linear relationship between them and SSI means that SSI is able to possibly create more separation in the feature space between the clusters. The fact that MI correlations are higher for SSI than the Pearson correlation does in fact imply that the relationship to e.g. wind speed is non-linear. This is mentioned at the end of section 4.1.

Comment 12

Line 255: Why is it important to know that the method is 'heavily' used? This does not imply that it is the most appropriate method for this study.

Response:

The word has been deleted from the text as it was deemed unnecessary.

Comment 13

Lines 281: What is the silhouette score?

Response:

The silhouette score measures the proximity of samples in other clusters to a sample in a specific cluster in the feature space. The silhouette score is the better (closer to 1) the closer a sample is to its own cluster's centroid than the nearest cluster's centroid that the sample is not a part of. Essentially, it tells how distinct the clusters are from one another. See Shahapure and Nicholas (2020) for further explanation. This reference has been added to the text as well.

Comment 14

Figure 2 caption: The caption should refer to table 1 for details of the intensity measures.

Response:

The figure caption has been modified to include this reference.

Comment 15

Line 365: Why do the SSI measures have no weight? What is the interpretation of this result?

Response:

The SSI measures have no weight in the Sparse PCA because they have very small variances compared to the other measures. Many ETCs have very small or zero SSI and only some have moderate or large SSI (the distribution is far from Gaussian even on a base-10 logarithmic scale). This causes the SSI variances to be much smaller than those of other intensity measures. The interpretation is that the SSI measures do not represent the variability of the whole dataset very well.

Comment 16

Line 395: Here the 4 clusters and their names are introduced. Are the clusters identified from the sPCA figure (fig 5)? Or have I misunderstood the methodology here? Also, since these names are used frequently in the remainder of the paper, I would suggest using bullet points so that they stand out in the text.

Response:

The clusters are not identified from the sPCA figure. Instead, they are identified by using the Gaussian mixture modelling (GMM) method which is introduced in Section

3.3. The GMM is performed with the identified intensity measures, which come mainly from the sPCA, as input. The use of the methodology has been clarified in the text by reformulationg the sentence to

> The cluster analysis was performed using the method described in Sect. 3.3 with the reduced set of intensity measures identified in Sect. 4.2 as input.

See also General Comment 1 of Reviewer 2.

We have changed the listing of the clusters to include bullet points. Thank you for the suggestion.

Comment 17

Figure 5: The words Calm/Windy, Dry/Rainy and Small/Big should be referred to in the figure caption.

Response:

The figure caption has been modified to include the sentence

> The labels Calm & Windy, Dry & Rainy, and Small & Big refer to the qualitative interpretation of PC1, PC2, and PC3, respectively.

Comment 18

Line 526: It is a bit confusing to use intensity here, since one of the clusters is also called intense.

Response:

We have not modified this sentence as throughout the manuscript we have discussed and referred to the intensity of extratropical cyclones and therefore do not think this is confusing. Additionally, we want to keep our terminology consistent.

Comment 19

Line 534 and 539: How are the cyclones in the XWS storm catalogue identified. If they use SSI then it is not surprising that a large number of the SSI cluster are contained in the XWS storm catalogue. Similarly, the named storms are those that lead to impact, hence they are biased towards landfalling storms. Can these impact-based metrics be used independently to verify the usefulness of the storm clustering technique?

Response:

The named storms in the XWS catalogue were chosen based on the amount of insured loss they caused. While one aim for the creators of the catalogue was to find an index which would rank the named storms highly and this index is similar to the SSI we used, the storms were not selected by using this index. Our result is consistent with theirs in this regard: SSI ranks these named storms which caused large insured losses highly.

We have not tested whether the impact-relevant metrics can be used independently to verify the usefulness of the cluster analysis. Of the impact-relevant metrics, SSI seems to have a large effect on the cluster to which storms get assigned. We have however shown this for only a small sample of high-impact storms so the result cannot be generalized. Furthermore, all investigated storms in the HighSSI cluster do not have large values in all of the impact-relevant metrics. For example, storm Christian/St. Jude belongs to cluster HighSSI but has below average values in PC2 and PC3, i.e. precipitation and wind footprint. However, storm Christian has above average wind speeds (not impact-relevant by our definition) which affect the determination of its cluster as well.

Comment 20

Lines 585-595: This is interesting information, but the authors make no link to the results in these studies to their study, so I'm not sure why this information is included?

Response:

We have extended the discussion of this result from

> Their northward-oriented tracks were mostly found near the eastern coast of North America, which means it could contain many of the same tracks as in our clusters HighSSI or Intense (e.g., post-tropical cyclones). In terms of MSLP, their northward-oriented ETCs were the most intense but the northeastward-oriented ETCs were faster-moving, while they found no significant differences in ETC lifetime between any of the clusters.

to

> Their northward-oriented cluster tracks were mostly found near the eastern coast of North America and were among the most intense ETCs in terms of MSLP. This indicates that the northward-oriented tracks could contain many of the same tracks as our clusters HighSSI or Intense (e.g., post-tropical cyclones). They also found that the northeastward-oriented ETCs were the fastest-moving, while they found no significant differences in ETC lifetime between any of the clusters. This result is different from ours, as we found a link between average dynamical intensity and ETC speed and lifetime.

Comment 21

Line 620-634: As above, it would be helpful if the authors could highlight the novelty of their results and how they build on the work in the previous studies described in this section. How is the cluster analysis a 'new perspective to the classification of ETC life cycles'? Could the authors be clear about what they are adding to the scientific literature?

Response:

The paragraph has been revised to be more clear about the stated claims and now reads:

> While these types of analyses are suitable for studying the precursors and forcing mechanisms of ETCs, we demonstrate that classification of ETCs based on their intensity benefits from an added level of objectivity via the cluster analysis. This can be seen in the overlap between the intensity measure distributions for different clusters in Fig. 6. Despite this overlap introduced by the objective method, our clusters can be at least qualitatively linked to classes of ETCs obtained with more subjective methods as described above. In fact, a possible course of future study is the identification of the variability in the ETC precursors and forcing mechanisms within our clusters. We believe that this form of analysis, which links the intensity and relevance for impacts to the genesis environment of ETCs, would offer a new perspective to the classification of ETC life cycles and possibly improve the predictability of ETC intensity.

Comment 22

Line 643: Given these limitations, could different criteria be used to identify Mediterranean cyclones?

Response:

We agree that different criteria could be used to identify Mediterranean cyclones better. However, this would likely introduce inconsistency and more subjectivity compared to the Atlantic cyclones identified with the criteria. Although we acknowledge the limitations in the criteria used, we show that the number of tracked Mediterranean cyclones is not very different from literature.

Typographical errors:

Comment 1

Line 166: Missing space before the bracket.

Response:

A space has been added before the bracket.

Comment 2

Line 167: Extra space after 2003.

Response:

The empty space has been removed.

Comment 3

Line 249: 'drawback' should be 'drawbacks'.

Response:

The word has been changed accordingly.

References

Joona Cornér, Clément Bouvier, Benjamin Doiteau, Florian Pantillon, and Victoria A. Sinclair. *Classification of extratropical cyclones using multiple measures of intensity: Data and Python code*, September 2024. URL `https://doi.org/10.5281/zenodo.11384417`.

ECMWF. *IFS Documentation CY41R2 - Part IV: Physical Processes*. Number 4. ECMWF, 2016 2016. doi: 10.21957/tr5rv27xu. URL `https://www.ecmwf.int/node/16648`.

Ting-Chen Chen, François Collet, and Alejandro Di Luca. Evaluation of ERA5 precipitation and 10-m wind speed associated with extratropical cyclones using station data over North America. 2024. doi: 10.1002/joc.8339.

Lorenzo Minola, Fuqing Zhang, Cesar Azorin-Molina, AA Safaei Pirooz, RGJ Flay, Hans Hersbach, and Deliang Chen. Near-surface mean and gust wind speeds in ERA5 across Sweden: towards an improved gust parametrization. 55(3-4):887–907, 2020. doi: 10.1007/s00382-020-05302-6.

Peter Bechtold and Jean-Raymond Bidlot. Parametrization of convective gusts, 2009 2009. URL `https://www.ecmwf.int/node/17487`.

Ketan Rajshekhar Shahapure and Charles Nicholas. Cluster quality analysis using silhouette score. In *2020 IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA)*, pages 747–748, 2020. doi: 10.1109/DSAA49011.2020.00096.