#2 - Reviewer

General comments

The paper by Sanuy et al. "Synoptic weather patterns conducive to compound extreme rainfall-wave events in the NW Mediterranean" presents (1) a synoptic-climatological assessment of compound extreme events over the NE part of Spain coupled with (2) the use of Bayesian networks that quantify the nonlinear links between the synoptic types and various variables describing the extreme events.

Overall, I find the study interesting and potentially worthy of publication, but only once the authors have successfully addressed the major issues. These issues comprise the methodology, the balance between the weather typing and its subsequent application, and the clarity of presentation (both text and figures).
We appreciate the comments and time dedicated to the review and are very grateful for the accuracy of the observations. The main issues pointed out by the reviewer will be addressed in the updated version of the manuscript.

After considering the feedback from both reviewers, we are updating the manuscript. The revised version will now include an analysis of the model skill based on various factors: (i) different combinations of variables; (ii) domains; and (iii) number of clusters of the KMEANs approach "N". Using the BN model skill results (synoptic skill) as a criterion, we will select the appropriate combination of domain, variables and N. We have nearly completed these tests, and preliminary results indicate that the number of clusters can be reduced for all types of events without a significant loss of model skill.

We will enhance the manuscript by emphasizing the BN results, both in methodology and results section. As a result, the description of weather types will be condensed and simplified. Additionally, we will improve the quality of figures, add corresponding label, and enhance figure captions in both the main manuscript and supplementary material.

Specific comments

In synoptic climatology, it is well known that links between synoptic-scale circulation and any conditioned surface variable are sensitive to how the circulation domain is defined, both in terms of its size and localization relative to one's area of interest. Based on the presented results, I do not understand why for studying extremes mainly dependent on close-by lows or troughs the authors decided to analyse atmospheric circulation over such a broad area. This is particularly striking for precipitation variables, which require smaller domains for good skill (Beck et al. 2016; IJC 36:7). This issue demonstrates itself when the authors train their networks, the skill of which is very low for precipitation. The authors suggest that including other variables including large-scale teleconnections may help, but I suggest that their primary focus be on smaller rather than larger scales. I strongly recommend that the authors experiment with the size and location (i.e. centre versus off-centre) of their circulation domain and assess the sensitivity of the networks' skill to these changes. If classifications are trained independently on each type of event, there is even no reason to use an identical region for each of them.

On a similar note, I am not convinced that including MSLP, Z500 and 10m wind components adds in this particular case any synoptic skill. The authors claim they were motivated by Miró et al. (2018), who however did not analyse events related to this analysis, as the authors claim in 344, but rather cold-air pools in which decoupling between local (site) and regional circulation systems was

crucial. As part of discussion, the authors should include a sensitivity study showing how their BN outputs respond to inclusion of multiple mutually strongly correlated circulation variables.

The revised manuscript will feature an analysis of model skill based on different variable combinations, domains and number of clusters of the KMEANs approach "N".

The BN model skill, serving as a proxy for synoptic skill, will guide the selection of a suitable combination of domain, variables and N. Preliminary results indicate an improvement in model skill for some of the reduced domains, but not for all event types.

The analysis of atmospheric variables reveals that a chosen set (mslp, z500, u10, v10) demonstrates robust model skill across event types and interest variables at the local scale (both Hs, and P24h, for compound or uni-variate extremes) in some domains, although (mslp, z500, u) is also showing good performance. Depending on the results of the remaining tests, the chosen variable set will also be used in the updated version of the manuscript to classify the extreme events, or changed by the better alternative.

Based on the description, it is not clear whether the circulation variables were standardized prior to PCA decomposition – the authors only mention they used anomalies, which would not be enough to account for the differences in variables' variance. In such a case, only one variable would have a dominant impact on the classification output anyway.

Variables were standardized. This will be indicated in the updated manuscript.

This relates to my other comment. I do not understand why were the weather types defined independently for each of the extremes' types, why the authors decided to define that many types (I do not claim that it is wrong but – again – no evaluation of the effect on the networks' skill was carried out), especially considering the fact that each classification was subsequently (manually) clustered into three "supertypes" that have long been known to link to the studied extremes in the region.

In the study, classifications of all extreme days were initially tested before trying independent classifications for each type of the extreme.. The obtained skills were found to be higher when dividing the data per event type, resulting in a lower total number of weather types.

In the updated version of the manuscript, the BN-model skill will be used to determine the optimal number of clusters, comparing various classification domains and variables.

For compound and wave-only events, N values of 6,10 and 14 are under exploration, while for rain-only events, N values of 10, 18 and 26 will be considered. These different N values correspond to the varying number of each event type.

Compared with the synoptic types, the description of which seems unnecessarily too lengthy and overly detailed to me, the text describing the networks seems way too short. Note that the classification and its descriptive analysis represent an interesting exercise. However, the added value consists (or, should consist) in the subsequent analysis.

The BN skill will serve as a metric to guide the selection of variables, the number of clusters, and domains. Subsequently, it will be used to evaluate the predictive potential of different accompanying variables and describe the final classifications across the territory. In the updated manuscript, greater emphasis will be placed on the BN part of the method and results, while reducing the description of the synoptic types.

I strongly suggest decreasing the number of abbreviations used in the text. For instance, why using C/SR/SW instead of simply compound, rain and waves? In some parts, where these are combined with abbreviations of variables and types, the text is extremely hard to read. Last, please try to simplify your terminology, better explain it, and use it consistently.

This will be addressed in the new version of the manuscript.

The quality of the figures is not great. It is practically impossible to see the background maps on screen, let alone when printed.
Figure captions and quality will be improved (and corresponding labels will be added to the figure panels).

Technical corrections/queries

61 in what sense are PCA- or CA-based classifications "more subjective than those" mentioned above?
We actually meant to say "objective". This will be corrected in the updated manuscript.

71 One may argue that the study by Sanuy et al. (2021, HESS) already did this, albeit to a limited extent
The reference will be added to the introduction.

73 It is not clear what "different mechanisms" mean; different to what?
The authors concluded that the mechanisms in the NE Iberian Peninsula are different from those of the other sectors. This will be clearer in the updated manuscript.

74-76 Please reword this sentence, it is hard to understand
The sentence will be rephrased.

90 did you mean "forcing on"?
Yes, this will be corrected.

2.1 Consider moving 89-93 (local scale) after the description of the larger-scale factors of extreme events
This will be considered in the updated manuscript.

128-129 why was MSLP abbreviation defined in 82 but z500 was not?
Abbreviations will be consistently defined at first appearance.

136 Are waves also a meteorological driver or is it a typo?
Yes, we refer to meteo-ocean drivers. .

137 I do not understand what "weather classifications associated SWP" means. Aren't SWP a synonym for weather (circulation) classification/types?
We will rephrase the sentence to: "Weather classification: obtaining SWP for each event type"

138 Isn't the classification method used to identify dominant SWPs? Also, what is "dominant", "critical", "target variables", and "SWP system"? I like the inclusion of the general framework and Fig. 2, but at this point they are not very clear, i.a. due to inclusion of abbreviations that have not been defined
Both the figure caption and the manuscript text referring to the General Framework figure will be modified based on reviewers' comments.

144 Was P24h explained in Sect. 2?
The abbreviation will be introduced in section 2.2.1.

158 P24 and P24h: what is the difference?
It is the same, P24. We will check for consistent usage throughout the manuscript.

161 "as as"
The typo will be corrected.

174 Does "separately" mean that each variable's anomaly fields were decomposed separately?
Yes. This will be clearer in the updated version.

175 What do you mean by "fundamental" modes and how does they relate to the anomalies?
It was a way of describing what the PCA does with a given dataset. The phrase will be removed.

179 So how many PCs/clusters were finally selected?
In the updated manuscript, this section will be modified, and the Knee-test will no longer be used to determine the number of clusters. Instead, an analysis of the BN-skill will be utilized. The final classification will specify the number of principal components (PC) and clusters.

181 How were they grouped? You need to show the patterns and refer to them from here. Also this text suggests that you join the types in three final types for which you present results, which is not the case. Please reword. Furthermore, I recommend using a different term for your overarching three patterns, such as "supertypes", to clearly distinguish them from SWPs.
The "classification" in "supertypes" will undergo a deep review. Instead, the structures observed in the most severe weather types will be described (i.e. presence of Mediterranean Cyclones, presence of Cut-Offs, etc.).

197 all?
It is a typo. The sentence will be rephrased to "linear relationships between all variables"

200 what is parent?
In this context, synonym of predictor, and opposite to target. This will be clearer in the updated version.

201 is "SWP system's proficiency" the same as "classification's ability"? If the term system is important, please explain it sooner.
The term system is not important. We will use classification's ability.

204 correct "(SWPs)"
This will be corrected.

209 Why are not all of these parameters/variables mentioned in Sect.2?
Hs, P24, Tp and wave direction are presented in Sect.2 (they are raw data). Wave power content is calculated from Hs and Tp as explained in Section 3 (is one of the first methodological steps in event identification). The season of the events is also retained (this will be included in section 3, as it is not mentioned).

223 Standard deviation of the skill?
Standard deviation of the predictions.

230 Why do you repeatedly define abbreviations that you then don't even consistently use?
Abbreviations related to the event types will be dropped.

235 Please add panel captions into the figures, and refer directly to individual panels from the text as much as possible.

This will be done in the updated version of the manuscript.

252 How is there a decrease from North to South is the values are identical for the central and southern regions?

The sentence will be rephrased.

263 What does "their" refer to?

The sentence will be rephrased to "The weather types were linked to the event season"

266 What was verified?

The sentence will be rephrased.

266 The sentence "highlighting...severity" makes no sense

The sentence will be rephrased.

273 Are local multivariate events the same as compound events? I am becoming lost in the terminology.

A compound event may affect a basin as multivariate (both components exceed the extreme threshold in the same basin), while affecting other basins with only one of the components. Likewise, a compound event may not be multivariate at any basin, but compound on a regional scale. We will simplify the description of the weather types, focusing on the intensity and probability of local spatial co-occurrence. In any case, we will try to be clearer when talking about the different types of compound events to avoid confusion and "noise".

282 "The dominant..." > "The two dominant..." ... "all of them" means those in Figs 5+6 or something else? Please explain/reword "minimum relative cumulative spatial temporal all basins probability"

This will be rephrased in the updated manuscript.

285 What do you mean by "unstable conditions"?

This will be rephrased in the updated manuscript.

287 In all seven? I can hardly check because of the quality of the map, but it seems unlikely that all seven types (than even do not belong to the same "supertype") have the same feature.

All seven types present wind fields over the Mediterranean characterized by long trajectories over the water from E to W.

297 Please explain the abbreviations of types in the text, or use generic names (e.g., type1, etc.). It is impossible to remember which type is which.

Abbreviations will be dropped, as the description of weather types will be simplified.

346 There are more objective alternatives to visual checking for similarity. One can calculate a pattern-correlation matrix, or a distance matrix, project all SWPs to the first principal component plane, or alternatively one can use e.g. Sammon mapping to test whether SPWs that you identify with different supertypes truly occupy distinct parts of the data space without the linear/orthogonal PCA constraints.

The "classification" in "supertypes" will undergo a deep review. Instead, the structures observed in the most severe weather types will be described (i.e. presence of Mediterranean Cyclones, presence of Cut-Offs, etc.).

370 is it the same skill as in 221? I suggest adding a reference to the section where it is explained
Yes. This will be done in the new version

393 remove comma before "leading". Better explain "dominant" and "general"
This will be addressed in the updated version.

424, 737 etc. What is "weather configuration"? Please select a clear terminology and use it consistently.
This will be addressed in the updated version.

429 This describes the results of the reference, or yours?
Ours. The discussion section will undergo a deep review, adapted to the new version of the manuscript, and all minor comments will be addressed.

448 consider changing "this work" to "their work" or similar
The discussion section will undergo a deep review, adapted to the new version of the manuscript, and all minor comments will be addressed.