Review of wcd-2023-12

The seasonal forecast skill of European windstorms is investigated in this study. In particular, the dynamical factors that potentially drive the known skill in seasonal forecasts of storms are assessed. It is shown that for four key dynamical drivers of cyclones: their representation in the seasonal forecast model is similar to ERA5, the seasonal forecast of storms is correlated to the dynamical factors in various upstream regions, and well forecast storm seasons are correlated with well forecast dynamical factors. This topic is definitely of interest to the community and fits well within the scope of the journal. The methods used seem appropriate (though they are not always well explained) and the results are interesting. However, I found the paper to be very hard to follow. There are many poorly-worded sentences and poorly-described figures. I would think the manuscript would be suitable for publication after a thorough proof read and strong edit for clarity and completeness. More in depth comments are included below.

[LD] Thank you for the comment and following suggestions.

Major comments:

1. Clarity of writing.

I found much of the text hard to follow. The article would benefit from a thorough rewrite to draw out the main aims, results and implications of the study, which are currently being lost within the somewhat unclear text/structure and confusingly worded/long sentences. For example, the introduction contains several paragraphs about seasonal forecasts and the dynamics of extratropical cyclones, but they are not well linked with each other or related to the aims of the study (which are not really mentioned until the end of the introduction). I have listed some sentences that were unclear to me below, but it is not an exhaustive list and I recommend the entire manuscript be checked for clarity. Also, several sections begin with a question, which is presumably the question that the section aims to address. The questions need to be properly introduced and answered if they are to be included, it reads as draft-like in its current format.

[LD] The manuscript can be checked in total for more clarity with a focus on clear and shorter sentences, more description in Figure captions.

The methods need to be more clearly explained as well. In particular, it is not clear what exactly is being shown in the Figures or how it is calculated. Most of the results show correlations but there is no information on exactly what is being correlated. The results certainly could not be reproduced with the information that is currently included.

[LD] Correlations are always Kendall correlations which can be stated in the method section, and the figure captions can be adjusted with more information to clarify what is correlated in each figure.

Unclear sentences:
L7: "Following Glosea5 factors' validation contributing to windstorms"
[LD] This part of the abstract was suggested to rephrase by Reviewer I as well and can be clarified with numbering the investigation steps: "Following GloSea5 factors' are (1) validated on the physical connections to windstorms, (2) investigated on the seasonal forecast skill of the factors themselves, and (3) assessed on the relevance and influence of their forecast quality to windstorm forecast quality."
L23: "Windstorms in this study are thus more related to the direct impacts of a cyclonic system". More related than what?
[LD] Further details could clarify this sentence "Windstorms in this study are thus more related to the direct impacts of a cyclonic system rather than just the low-pressure systems."
L59: "Hence, it is connected to cyclonic systems and can be an indicator for their strength and location over the North Atlantic".
[LD] This sentence can be adjusted and clarified.
L190: "but the time coherent link between storms and factors is also of great interest, hence a correlation analysis between the factors' time development and windstorm frequency is used for validation"
[LD] This sentence is explaining Fig. 4. We could write it in shorter sentences to make it clearer. "Composites are categorical separations of data sets, which are useful for identifying the difference between two data sub-samples clearly. A time-coherent link between storms and factors is also of great interest. Hence, a correlation analysis between the factors' time development (as time series) and windstorm frequency (as storm counts) is used for additional validation (see Fig. 4)."
L204: "After knowing that relevant factors are well represented in their connection to windstorms not only from an ensemble mean perspective, but also within individual ensemble members and thus representing a consistent physical development, the next step tests if these factors themselves are well predicted."
[LD] Here as well, with writing it in shorter sentences, we can try to make it clearer. "The previous results summarise that relevant factors are well represented in their connection to windstorms. This had been shown for an ensemble mean perspective (with composites, Fig. 3) but also within individual ensemble members (correlations per member, Fig. 4). Thus, the GloSea5 model represents a consistent physical development between respective factors and windstorms. The next step tests if these factors themselves are well predicted."
L207: "Thus, in those regions of important connections between factors and windstorms (section 4.1) they should be well predicted to make an influence for the windstorm forecast performance."
[LD] Maybe reduce the sentence to make it clearer: "The storm-relevant regions (section 4.1) should be well predicted to have a positive influence on the windstorm forecast performance."
L306: "With mostly agreeing physical connection between windstorms and individual factors within the observational and model data these connections may enhance model forecast performance when the individual factors are well forecast themselves".

[LD] We see that is was a confusing sentence. By breaking it down into 2 shorter sentences, we hope the message gets clear: "The physical connections between windstorms and individual factors within the model data mostly agree with the connections in the observational data. These connections may enhance model forecast performance when the individual factors are well forecasted in the storm-relevant regions."

L323: "For all four factors the model provides positive forecast skill within relevant regions, means the model performance for the individual factor is positive and well predicted seasons in these regions, supporting skilful windstorm forecasts."

[LD] Same as above, we hope a paragraph with shorter sentence will make this statement clear: "The model provides positive forecast skill within relevant regions for all four factors, which means the model performance for the individual factor is positive. The final investigation step shows that well-predicted seasons of the factors in the relevant regions support skilful windstorm forecasts."

L333: "A similar scattered result is resulting for all approach steps for the SST gradients."

[LD] We think the double result is not well chosen, so we could change to "result is seen for all"

L344: "which give new knowledge where the windstorms forecast skill might originate and where additional efforts, beside the also for windstorms existing signal-to-noise paradox"

[LD] Hopefully a rephrasing along these lines help for clarify ", which implies new knowledge about where the windstorm forecast skill might originate. This also reveals areas for additional efforts needed to potentially improve windstorm forecast skill over the downstream end of the North-Atlantic storm track, alongside the also for windstorms existing signal-to-noise paradox"

2. The dynamical factors.

Much of the analysis focuses on four of the dynamical factors that are deemed most influential for cyclone development, yet there are 20 (by my count) that are included in Table 1 and Figure 1. I wonder if it is necessary to include all the factors in Table 1 and Figure 1 as you do not really mention them in the text (the coloured boxes in Figure 1 are not defined either).

[LD] Figure 1 will be adjusted with a legend.

The schematic in Figure 2 is also not properly described. I would recommend removing the Figures and Table and simply listing the predictors you chose to analyse in the study. If you do keep all the predictors in the manuscript then there should be a much more thorough description of what each means and how they relate to cyclone development (though I'm not sure what the point of this would be as the majority of the predictors are not included in the main text).

[LD] We understand this comment, but wanted to show, that we have not only checked the four focused factors but a bigger list of potential factors. Table 1 could be moved to the appendix or reduced to only the focused factors. Fig. 1 and 2 are

supposed to show the different levels of interaction between factors and cyclones/windstorms. We think about removing Figure 1 from the manuscript and move Figure 2 into the appendix, so reduce complexity of this part. But still more details can be explained in the text.

There is also no clear explanation on how the four included predictors are chosen (you say they "highlight the postulated link to winter storms clearly and best"). What metric is used to determine this? This information would potentially be more beneficial to show than the schematics.

[LD] We choose to have 2 primary and 2 secondary factors in the paper, but more in the appendix, to not overload the manuscript. The way of choosing was a step-by-step investigation which factor show a clear result throughout all investigation steps.

3. Selection of good and bad forecasts.

I am somewhat confused on how you separate good and bad forecasts for the results presented in Figure 6. In section 3.3 it says you separate forecast years into good and bad by comparing their storm counts to that in ERA. But then in section 4.3 it says you separate them into good and bad by considering the skill of the forecast factors (though it is not clear exactly what you mean by this). I have a number of concerns about the approach regardless:

-Are you considering at all the temporal aspect of forecast skill or if the skill is actually related to wind storms? If you are just comparing the mean values of the factors in the different regions across the entire forecast then I'm not sure you can relate this purely to windstorms. For example, you might have a low value of MSLP gradient in the different regions that is well predicted and which is associated with a good prediction of a reduced number of storms. Therefore the skill may increase over the UK but not in relation to storms. (I could be misunderstanding what is shown in the plot.)

[LD] I am not fully sure I understand this comment. We are only using the factor forecast skill to separate the seasons. And then investigate if these different sub-samples of season have different characteristics in storm forecast skill. We believe there are more ways of doing this but this would exceed the aim of the manuscript.

-To me, a more intuitive approach would be to consider the factor skill in the regions when a storm is identified. Then you could show that when a storm is in the forecast and the factor regions are well predicted, the storm is well predicted over the British Isles, and vice versa. You have the tracks for the storms so this should be feasible.

[LD] This sounds like a reasonable approach, but is not the idea we wanted to follow with this study. Our aim wasn't to look at individual storm tracks or events, we wanted to look at a general state of the atmosphere during the windstorm seasons and depending if it is a stormy or not stormy season.

-Do you require the skill to be good in all the factor regions? If so, have you tested if a particular region is most important. I.e. does the forecast skill over the UK increase more if the factor is well predicted in a particular region?

[LD] the regions have been tested individually itself (see table in the appendix), the figures shown in Fig. 6 are only the ones with the highest change (between well and badly predicted factor) in correlation. Meaning if multiple boxes are shown all boxes are considered in this particular panel, but the rest has been tested as well, but was less significant.

-Is the difference in the left and right columns of figure 6 just that the regions used to define the good and bad forecast skill are different? But the method is the same apart from that?

[LD] yes exactly, we could add something like "…separation by the Factor-skill-view (left column) and the Process-based-view (right column). The separation is based on spatial averages over the shown boxes from Fig. 5 for the left column and Fig. 3 for the right column, …" in the Figure caption.

-Have you tested different metrics of forecast skill? The results presented may be sensitive to the metric you use to determine forecast skill (please state clearly what this is). It would be good to try other metrics and compare results.

[LD] This study is based on a previous study of the authors, where we show also different metrics for the forecast skill. Kendall correlation was the most intuitive to use and too understand, but we agree, that the results can be dependent on the chosen metrics.

Minor comments:

L45: the Eady Growth Rate parameter is not itself a source and intensifying factor for extratropical cyclones. Strong baroclinicity is (i.e. what high values of the EGR parameter represent). So this sentence needs rephrasing.

[LD] This sentence should be correct by changing and to which "The Eady Growth Rate (EGR) parameter (Eady, 1949) is used as a standard measure for baroclinic instability of the atmospheric flow ~~and~~ which is known as a source and intensifying factor for extra-tropical cyclones (Hoskins and Valdes, 1990)."

L50: "These variables were also used in other studies". Other studies about what? How were the variables used? Some additional context is needed here.

[LD] We will write more context to these studies.

L77: "This could lead". This seems vague and weak. You could say something like "the aim of this study is to better understand…". Or something similar.

[LD] Thanks for this suggestions, something along these lines can be used to change the sentence

L115: Do you mean local PV? Remote PV anomalies can influence cyclones via action at a distance.

[LD] We are not specify local or remote, as some investigation/method steps consider a spatial distance of the factor to the target storms (e.g. Fig. 3, Fig. 4 or Fig.

6). They have the target of UK storms but investigate the factor (here PV) for the whole North Atlantic.

L118: Did you test if your results are sensitive to the averaging length? 3 months seems quite a long time period to average for dynamical factors relating to cyclones.

[LD] We agree with that. The aim of the study was to look into the seasonal time scale. Some factors (EGR & PV) have been tested as bandpass filtered version to take into account that they are important on a smaller time scale, but their results were very scattered and not conclusive. This is included in the Method-section in the appendix.

L150: Please define what is meant by tau_b Kendall correlations. How are they calculated?

[LD] This can be clarified with the following sentence "Kendall correlation is a similar measure to the commonly used Person's correlation but investigates ranked time series and is less subject to normally distributed data." We could rather not add too many equations of established statistics, they can be found in the mentioned citation.

Throughout: use of chapter instead of section!

[LD] I hope this reviewer means the same as the first reviewer. We decided to go with section for the whole manuscript and changed it to be consistent.

Figure 3: Are there compensating errors here? I.e. do the strong storm seasons look similar in GloSea and ERA, as well as the weak storm seasons, their differences might look similar for the wrong reasons.

[LD] That is a good point, this can be checked internally and mentioned in the manuscript.

L187: would a dipole suggest a shift in precipitation location rather than an overall increase?

[LD] That is true, thanks for the comment, we will change the "more precipitation" to "shifted precipitation"

Figure 4: unclear exactly what you are correlating here? You correlate the number of storms with what metric of the dynamical factor (mean in the boxes?).

[LD] Caption can be adjusted with an additional explanation like "Correlation Maps between seasonal storm counts over the UK and dynamical factors (averaged in 10x10∘ regions)"

Figure 4: can you include the correlation values for GloSea as well? This would allow for easier comparison than comparing redness/blueness. The histograms are very small here as well.

[LD] The value inside the ERA5 row is not the correlation value.

Figure 5: again, it is not clear what is being correlated here.

[LD] Can be adjusted by something like "Kendall Correlation maps for selected dynamical factors between ERA5 and GloSea5 per grid cell"

L231: what aspect of factor skill are you referring to here? The mean value of the factor in the region? Temporal evolution? Please state explicitly.

[LD] This is the explanation of one approach for the investigation step #3. This sentence is about the regions selected to create subsamples of the storm data. By

adding a few more words, we hope this sentence is clearer, like "that show coherent regions of skilful forecasts for the individual factors".

L276: the aim of the study here should be more clearly stated in the introduction [LD] This has been stated in the introduction, but we could extend the introduction sentence like this "These skilful storm forecasts found in seasonal hindcasts lead to the motivation for this study. This study aims to understand which dynamical factors drive the seasonal winter windstorm prediction skill, whether as primary or secondary related factors."

Technical corrections:

L2: the seasonal forecast of —> seasonal forecasts of

L5: I'm not sure if ERA5 and GloSea5 should be included in the abstract without defining them. Perhaps change to "a reanalysis product and a seasonal forecast system".

L10: What three steps? [LD] by adjusting the abstract we added numbers for the 3 investigation-steps

L21: use rare or extreme. Do not need both.

L26: remove "from" before "different regions and hazards".

L44: "investigated" —> "have investigated".

L46: I'm not sure if i.a. is right here?

L54: Need to define theta_e before you use it.

L73: "Further on". Further on than what? The study you refer to is from 2015 which is earlier than those mentioned previously. [LD] rephrased with "another factor discovered by …"

L90: GloSea5 is defined earlier, though not fully? [LD] Added earlier around line 35

L109: in —> is

L127: "exemplary" means very good. I do not think that is what you mean here. [LD] I am not a native speaker, hence, I have to google and trust my online dictionary, but I find "exemplary" as the adjective for an example

L143: bad —> badly

L155: up on —> upon

L174: less strong —> stronger? [LD] less strong = weaker, have changed this

L189: is —> are

L198: outside —> upstream? [LD] no "outside", not only upstream but everywhere where it is not storm-relevant

Figure 4 caption: column —> row

L206: using "Thus" to start two sentences in a row, should be changed.

L212: upstream —> downstream? [LD] no, upstream is with the flow, meaning east of something in the midlatitudes

L213: downstream —> upstream? [LD] no, same reason as previous comment, vice versa.

L235: Does —> do

L235: remove "would" after storm.

L264: might be theta_e —> which might be theta_e

L265: is SST an atmospheric state? [LD] changed to "global"

L328: which —> who

[LD] Thanks for the small corrections, will be changed.