Review of egusphere-2022-288
Anonymous Referee #1

Referee comment on "Skillful Decadal Prediction of German Bight Storm Activity" by Daniel Krieger et al., EGUsphere, https://doi.org/10.5194/egusphere-2022-288-RC1, 2022

Review of the paper Skillful Decadal Predictions of German Bight Storm Activity

The paper investigates the decadal prediction skill for MSLP anomalies and German Bight Storm Activity in a 64-member ensemble of yearly initialized decadal hindcast simulations. These are based on MPI-ESM-LR and cover the period 1960-2018. The authors use both deterministic and probabilistic skill metrics to analyse the prediction skill for different lead times.

The paper covers an interesting and relevant topic that could be of interest to NHESS readers. It is clearly structured and the presented topics are covered with sufficient material. However, some aspects in terms of methods, analyses and results are unclear and/or not well described/discussed. Here the manuscript needs some general improvements in order to enhance its clarity and readability. Please refer to the main points and specific comments below. If these points are taken into account, I think the article would be a useful addition to the literature.

Main points:

- DATA
  - Just to clarify, you are not using the MiKlip data, but have constructed your own decadal prediction system? I was not sure until I got to line 102...
  - I do not quite understand how you constructed the 64-member ensemble (L104-111). Please describe this in more detail.
Please clarify which decadal runs you chose. If you are looking at the period 1961-2018, did you select all runs that include those years regardless of the lead time, or is the last run you selected the one that was initialized in 2008?

**METHODS**

- **Lead times, part 1:** The selection of lead times seems somewhat arbitrary. Why did you choose 4-10 and 7 and not 1-7 and 4 or 2-8 and 5 ...? Have you checked whether your results/conclusions would be different with a different choice of lead time?
- **Lead times, part 2:** In L126ff, you state that you focus on lead years 4-10 and 7. However, this only applies to the MSLP anomalies, since you show all possible lead year ranges for GBSA. Please be more specific in this regard.
- **Pressure reduction:** Is this a standard procedure for calculating MSLP from modelled surface pressure? Could you add a reference for equation 1? Does it affect the comparability of your results if you use direct MSLP for one half of the ensemble and calculate MSLP for the other half?
- **Region of interest:** Please clarify that you are analysing MSLP anomalies for the entire North Atlantic basin (including the German Bight), whereas the GBSA analyses focus only on the German Bight.
- **Selection of grid points (L140-144):** This information refers to the generation of GBSA time series, correct? If so, either integrate it in the respective paragraph (L146ff) or clarify why you need to select three grid points. At the moment, the whole paragraph comes a bit out of nowhere, without a clear link to the preceding/subsequent paragraphs...
- **Generation of GBSA time series:** Did I understand correctly that the time series cover the whole period 1960-2018, while you only use the period 1961-2010 for the standardization?
- **Prediction skill:** Please add a short explanation of why it is important to consider both deterministic and probabilistic skill scores when assessing the skill of a decadal prediction system.
- **ACC:** Although this should be common knowledge, please add the possible range of ACC and an explanation of what the different values mean.
- **ACC versus BS:** Be careful when using f and o in equations 2 and 4. You chose the same letters, but they have different meanings (value for ACC, probability for BS). Consider replacing f and o in equation 4 with capital letters.
- **Choice of BSS:** Out of curiosity – why did you choose the BSS rather than the ranked probability skill score (RPSS)? Since you are interested in three categories (low/normal/high), the RPSS seems the more natural choice to me as it also contains some information about the distance between model and observations.

**RESULTS**

- **Some thoughts on L234-242:** Could it be that the initialisation has a “negative” impact in the first years (some kind of initialisation shock) – which would explain why the predictive skill is highest for lead time ranges starting in year 3 and 4? This would also fit (to some extent) to previous studies on wind-related variables like Kruschke et al. (2014) or Moemken et al. (2016). However, these studies use uninitialized historical simulations as reference and not persistence...
  For temperature, several studies show high predictive skill for later/longer lead times (e.g. Feldmann et al., 2019). This increase seems to originate mainly from the long-term climate trend. However, I have never heard of the importance of climate trend for decadal predictions of wind-based parameters...
- **L304-338:** These paragraphs seem to be more of a general discussion of your results and are not really related to the rest of section 3.2.2. Therefore, it might make sense to introduce a new section (3.3 Discussion) or new chapter (4. Discussion) for this part of the manuscript.
- **Persistence as reference:** Many studies dealing with decadal prediction systems use uninitialized historical simulations of the same model or simple climatology as reference. Is there any particular reason why you have not tried this as well? Please
do not get me wrong – I think it is a strength of your study that you consider persistence and random guessing. It just makes it harder to compare your results with other studies on decadal predictions.

- **FIGURES**
  - For readers unfamiliar with Germany (and the German Bight in particular), it might be helpful to include a figure showing the region of interest. In this, you could also mark the grid points given in Table 1.
  - Figure 2: Please add some explanation in the text (L226-230) about the structure of the plot (that it shows all possible lead time combinations etc.).
  - Consider simplifying the captions of Figures 5 and 6 (the same applies to B3 and B4) by saying something like “Same as Figure 4, but for ...”.

**Specific comments:**

- Moemken et al. (2020) should be Moemken et al. (2021)
- L14-16: Please add a reference for this statement.
- L17: “certain types of extremes” – Can you give one or two examples?
- L23: “potential” instead of “potential value”?
- L27: Please add some of the standard references for the topic of decadal predictions.
- L28: Please add explanation of the abbreviation “MiKlip”.
- L35ff: Moemken et al. (2021) have analysed a dynamically downscaled component of the MiKlip system – please add this information somewhere.
- L39-40: Consider citing Reyers et al. (2019) at the end of this statement.
- L49/50: “an assumption which was shown by Krueger et al. (2019) to be valid” – please rephrase.
- L62-65: Could you add a reference for this statement?
- L68: Probably a stupid question – but would it be better to explain what a hindcast is?
- L73: “skill of the model system” instead of “skill of model system”
- L76: Please add a reference for the Brier Skill Score.
- L85-86: The use of two parentheses directly following each other looks strange – can you merge them? The same applies to lines 94-96.
- L155: “model's prediction skill” instead of “model’s predictions skill”
- L168: “The BSS is defined as” instead of “The BSS defined as”
- L179: “calculated” instead of “caulated”
- L193: “prediction skill” instead of “predictions skill”
- Figure 1: “Deterministic prediction skill” instead of “Prediction skill”?
- L210/211: Consider placing “(Fig. 1b)” after “... ACC is lower for lead year 7”
- L358: “predictions for short lead year ranges” instead of “predictions short lead year ranges”
- L394: “in such a way that” instead of “in such a way so that”

**References**