EGUsphere, referee comment RC3
https://doi.org/10.5194/egusphere-2022-288-RC3, 2022
© Author(s) 2022. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on egusphere-2022-288

Anonymous Referee #3

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-RC3, 2022

Krieger et al have investigated the predictability of German Bight Storm Activity for various interannual lead times and averaging periods. They find limited skill, but some hope that high and low storm activity is predictable. It's a well written and clear paper and it is great that both deterministic and probabilistic skill is covered. I recommend that this paper be accepted, subject to some minor revisions as detailed below.

One issue is that the paper focusses on some unusual forecast lead times (4-10 years and 7 years) without properly motivating why they use these. It would not seem the most interesting lead times for a user of a storm activity forecast. There is frequent reference to short and long averaging periods and I was not always sure whether that referred specifically to these two periods or had been generalised somehow. But if it is the latter, it is not defined. The language needs to be cleaned up around this.

Figures which lead to firm conclusions are squirreled away in an appendix. I suggest all important figures need to be in the main paper. See minor comments below.

Another issue is that there is no deterministic skill for mslp anywhere near the German Bight (Figure 1), so how do you explain that you have skill in predicting storm activity there? This needs to be covered in the discussion.

Negative skill is presented as useful skill. It is true, you could multiply the forecast by -1 and get a good forecast on average. The problem is that the skill is possibly negative due to a poorly modelled teleconnection and if there is an individual year when that teleconnection is not strong, multiplying by -1 could be the wrong thing to do. Better to assume negative skill is not useful even if it is significant.

Finally, the text often refers to "tails" of the distribution and "extremes" when in fact the

data refers to anomalies exceeding 1 sigma, which is neither in the tail or an extreme. These words need to be removed from the text.

Minor comments:
l.9 "short lead years" is this years 4-10? If so, this should be explicit in the abstract, which needs to stand alone.

l.126 The choice of lead years used in the study is not well motivated. Why 4-10? As a user I would want to know what next year brings and how the next five years will look on average. Do you have a user that is interested in years 4-10?

l.149 Would you not get a better estimate of the mean and standard deviation if you calculated it once over all members, rather than each one separately? (Assuming the members are interchangeable).

l.162 As you are looking at a quantity that has decadal variability there is persistence from one year to the next. This means you do not have independent samples as assumed by the Fisher method. Perhaps you should try block bootstrapping or another method that does account for this? If not, what is the lead 1 autocorrelation? Perhaps a time series of derived storm activity should be included.

1.195 "activity of a number n of years" would be clearer if written "activity of n years"

l.205 "skill in a circular area west of the british isles" and l.211 "skill over scandinavia" - both these regions have negative skill, which means the model is doing something wrong and potentially not responding to teleconnections correctly.

l.224 "anyhow" this word is unnecessary for the sentence and a bit informal for a scientific text.

l.241 "convincing explanation" how about initialisation shock? Lead years 2,3 and 4 are poor and 1 probably less good than it should be, which points to initialisation shock.

l.246 "We expect" a shift in the tail is usually caused by a shift by the whole distribution (and hence the mean), so you don't need a large ensemble size to detect shifts in the tail (Eade et al, 2012). Maybe you mean changes in the shape of the tail? You look at 1 sigma, which I would not describe as being in the tail.

l.251: "short and long lead year ranges" what does this mean? When does a lead year range go from being short to long? If you mean year 7 and years 4-10, be explicit and write that.

l.260 "skill emerges over the German Bight" It would help the reader to mark the German Bight on Figure 3.

l.271 "Overall" you are generalising here to all short and long averaging periods, which you cannot do as you have only looked at one of each. You also have not defined what is short and what is long - is the boundary at an averaging range of 2 years? 3 years?

l.302 "we conclude" figures that lead to conclusions should not be in the extended materials. I suggest you put the coin flip reference figures in the paper and persistence figures in the appendices.

l.319 "anyhow" see comment for line 224. Also "overwhelmingly" there is no room for emotional reactions in a scientific text

l.323 "ill-suited" you know the probability of falling into the upper category, can you not use this as a reference forecast?

l.329-338 Caveats belong in the conclusions section, not the results section.

l.345 "Brier Score" should be "Brier Skill Score". In addition, you also tested it against a 50% forecast.

l.361 This is the conclusions section, you may speculate upon a cause

Figure 1: Please add something to the maps to show where the German Bight is, not everyone knows. You could also mark out your three points used for the geostrophic wind calculation (or a on a later figure if the resolution is too coarse here).

Eade, R., Hamilton, E., Smith, D. M., Graham, R. J., and Scaife, A. A.(2012), Forecasting the number of extreme daily events out to a decade ahead, *J. Geophys. Res.*, 117, D21110, doi:10.1029/2012JD018015.