Natural Hazards
and Earth System
Sciences

Open Access

EGU

Discussions

Interactive comment on "Revisiting the synoptic-scale predictability of severe European winter storms using ECMWF ensemble reforecasts" *by* Florian Pantillon et al.

Anonymous Referee #2

Received and published: 27 June 2017

The manuscript investigates the forecast skill of extreme storms (often called wind-storms) in the ECMWF 20-year reforecasts. The ECMWF 20-year reforecasts are found to be skillful and ensemble spread well calibrated up to lead times of 3-5 days. After this the skill drops; storms are found to move too slowly and do not capture the intensity of observed events as measure by a Storm Severity Index. No systematic links between storm properties (size, intensity, etc..) and forecast skill is found. Some skill beyond 3-5 days is found using EFI and SOT indices, suggesting some utility for windstorm warnings at these lead times.

The paper will be of interest to weather forecasting community as it contains some

new and interesting results. In general, the paper is clear in its approach and figures are clear. I have a couple of specific comments on the paper (below) which should be addressed. I'd consider these major revisions, although I don't think it would take much to address these comments. Subsequently, I'd recommend the paper for publication provided these comments are fully addressed.

Specific Comments

1. Novelty of the study: The paper seems incremental in terms of progressing this area of science, since a lot of what is said in this paper was covered by Froude et al (2007). It would be helpful in terms of highlighting the novelty of this particular study if a) the Froude et al 2007 paper is discussed in the introduction and b) that the novelty of this paper is discussed in the conclusions.

2. Page 8. Lines 19 to 33 and figure 6. Figure 6 is very useful as it gives another sense of the utility of the reforecasts. However, I don't agree with some of the statements here about the validity or not using an ensemble mean. The statements seem rather confused to me. For example, we could say that for your MSLP analysis in figure 3 we shall choose a threshold error of 10hPa to indicate a useful forecast, and therefore we shouldn't compute an ensemble mean for when the bias in the ensemble mean went above this. You'd agree that this would sound like a strange and arbitrary thing to do, but this is effectively what you're arguing in this piece of text. This strange argument should be removed. Furthermore, I find it difficult to see how your results make the results of Froude et al 2007 invalid (line 19) as they looked at a different dataset. Could you be clearer here what you mean?

3. Page 12. Line 9 and Figure 7a and 7b. "The predicted SSI is thus divided by a factor of 2 for ease of comparison unless stated otherwise" Have you done this for the plots in Figure 7? If so then you will need to redo the plots without this adjustment and revise the text. There's no justification for dividing one dataset by an arbitrary number to make it more comparable to the other. Furthermore, why are the SSI much

larger in the reforecasts compared to ERA-I? Further down the page you say (Line 22), "ERA-Interim may also contribute to the cases of overestimation by underestimating the actual SSI due to its limitation at representing the mesoscale structure of some storms." You will need to provide some evidence of this statement (e.g. a reference). How much is ERA-I underestimating the true SSI? If ERA-I is very wrong, why are you using it as your main evaluation dataset? You'll need to address these questions.

Technical Comments

Page 1

Line 5. "…storms are correctly predicted…" correctly would mean without any bias. Perhaps "well predicted" or "predicted with only small forecast errors" would be a better expression. Line 9. "However, a large variability is" should be "However, large variability is…" Line 10. "and does not appear…". What is it that does not appear? Do you mean the "…and the predictability of storms does not appear…"? Line 21 "…and of their forecast in numerical weather prediction systems." Perhaps could be better expressed as "…and of the ability of numerical weather prediction systems to forecast them." In addition, I don't disagree with the sentence but references need to be added.

Page 2

Line 24-Line 28 and Figure 1. The sentences and the reference to Figure 1 do not belong in the introduction. They should be moved to the methods section.

Page 4

Line 6. "In a second step, the mimima of MSLP are connected between subsequent model outputs every 6 h to form tracks, if their displacement velocity remains consistent in time." This second half of the sentence doesn't really make sense. Could you split the sentence and make clear what "their displacement velocity remains consistent in time" means? Line 8. "filtered to exclude storms with a weak Laplacian" Can you specify the threshold is?

Page 5

Line 24. Do you include SSI values over ocean in your European spatial average? If so this doesn't seem like a good idea – does it make a difference if you use land-only values of SSI? Line 31 "resoved" should be "resolved"

Page 7

Line 14 and line 24. "Dispersion" often has a very technical meaning. I think here what you mean is "variability". There are other examples of this in the manuscript that should be changed for readability.

Page 8.

Line 14. Rephrase "The motion of cyclones was also too slow in the forecast but their MSLP was too deep." As something like "The motion of cyclones was too slow in the forecasts. In addition, but the forecasted MSLP was too deep."

Page 9

Line 4. "...on a specific day but anywhere over central..." would be better expressed as "...on a specific day over central..." Line 7 to 8. "...the predicted distribution of SSI is overestimated overall." I don't know what this means - what is the predicted distribution, is it the reforecasts? If so state this explicitly. Also state explicitly what the predicted distribution is overestimated relative to. Line 12. Can you add some detail to explain how you select events for the 99th percentile of SSI? Line 26. "Early Warning". This terms means something very different in different contexts. In some contexts, early warning only means 1-2 day lead time. I'd suggest being specific here in terms of timescale and call this section "Potential for Early Warnings on 5-10 day timescales".

Page 10.

Line 3-10. The description of Brier Skill Score should really be in the methods section.

Page 11

Line 1. "This value is taken for consistency with the SSI." Can you say explicitly what this means? Line 19. "...the optimal thresholds need to be levelled up and..." What do you mean by levelled up? Do you mean increased?

Page 12

Line 2. Should be "...which was noted..." Line 27 "The ensemble average is unbiased until day 3 to predict the position and minimum MSLP of the storms on the day of maximum intensity.", would be better expressed as, "The ensemble average has small biases until day 3 in terms of predicting the position and minimum MSLP of the storms on the day of maximum intensity." Line 30. Should be "...ensemble members captures the actual storm..." Line 30. "This bias is accompanied by an increase in ensemble spread by a similar magnitude, which suggests that the ensemble is calibrated, but only a minority of ensemble members still captures the actual storm at lead times beyond 3–5 days. This questions the relevance of using the ensemble average at longer lead times. This differs from a classical situation of averaging the ensemble members to smooth the unresolved scales, as the variables of interests are objects here rather than continuous fields" This appears to be a different argument than from earlier, where arbitrary thresholds were used to determine whether the ensemble contained the storm or not. Can you comment on this?

Page 13

Line 17. "The EFI and SOT indices confirm the skill of the reforecast at predicting the area covered by strong wind gusts until day 10 for storms as for the whole dataset." You argued a few paragraphs ago that few of the ensemble members actually predicted storm beyond 3-5 days lead time. If that's the case, how can there be skill at the lead times of up to a week? This needs to be explained in the conclusions. Line 29. "The predictability of the severe storms investigated here may not be linked to common factors but rather be due to characteristics of the individual storms." You've just argued in the previous paragraph that you don't have enough data to make this statement! So

how can this statement also be true?

Figures

Table 1 "Some particularly high or low values are emphasized in bold." This is a confusing thing to do – either remove the bold numbers or decide on a sensible reason for using bold numbers.

Figure 3 and Figure 4. Font used in legends is too small and needs to be substantially larger to be readable.