

Interactive comment on “Improvement in the decadal prediction skill of the northern hemisphere extra-tropical winter circulation through increased model resolution” by Mareike Schuster et al.

Anonymous Referee #2

Received and published: 3 June 2019

General comment:

The presented manuscript by Schuster et al. addresses the scientific relevant question as to how spatial resolution of a specific decadal prediction system influences the forecast quality of North Atlantic circulation measures. It is well within the scope of ESD and will find interest from readers in the research areas of decadal prediction, forecast quality assessment, climate modeling, and extra-tropical atmospheric dynamics. The study is well structured and offers novel results regarding the effect of spatial resolution in a decadal prediction framework for measures relevant to society and economy.

There are however some shortcomings in the argumentation supporting the conclusion likely due to an incomplete or imprecise description of the applied methods. The analyses provide insufficient explanation for some of the key results. The clarity of the manuscript will benefit from a revision of the writing.

Recommendation:

Re-Submission after few additional analyses and clarification of some results and applied methods

Specific comments:

1) The applied methods are often not clear. The use of an “evaluation software” is mentioned (P5L3). What does it actually do? When is the ensemble mean calculated, e.g. are the shown correlation maps means of correlations or correlations between ensemble mean and reference. Please provide clarification and add the applied calculation methods/equations. Could be as appendix/supplement.

2) The study suggests a direct relation between the mean bias of the ensemble mean and the anomaly correlation of the ensemble mean to the reference for one and the same diagnostic. The correlation is insensitive to the mean bias on grid cell level, hence anomaly. It appears large parts of the result section and conclusions are based on the assumption that a reduction of mean error/bias leads to higher anomaly correlations for the same analyzed quantity. This has to be revised substantially.

3) The hindcasts are presumably not post processed, e.g. corrected for time-varying bias, trend-adjusted, etc? Please clarify and state why this might be not necessary. Why is the approach of correcting biases of this study different to Kruschke et al?

4) Spatial resolution has been discussed to be a serious limiting factor to correctly reproduce climate mean state and variability in the context of decadal prediction (e.g. Hewitt et al. BAMS 2017, Smith et al. QJRMS 2016). This should be mentioned more prominently in the motivation and put in context of this study in the discussion.

[Printer-friendly version](#)[Discussion paper](#)

There are numerous studies about the effect of resolution in climate models in general including the effect on North Atlantic circulation measures (e.g. Davini et al. J ADV MODEL EARTH SY, 2017). How do they compare to this study?

5) The difference in mean bias for LR and HR in cyclone frequency is striking and given the very small differences in stormtrack activity somewhat unexpected, e.g. at 30W, 50N (Fig 2a vs Fig 3a). The result is apparently similar to Kruschke et al 2014. Kruschke et al compare uninitialized experiments in LR to NOAA's 20th Century Reanalysis. In their study the mean bias is up to 25 systems per winter over the North Atlantic and they mention a possible underestimation of cyclone frequency of the reanalysis. This seems at odds with what is shown here: A mean bias of up to 80 systems and more per winter in comparison to a different reanalysis product. Please discuss this Is it possible to estimate how much is due to the applied tracking method? One suggestion could be to interpolate the HR hindcast to the lower resolution and repeat the analysis. Will that change the results? This could be done for a single member and put as appendix. It is mentioned that LR overestimates weak and moderate systems. Why?

6) The ensemble spread is unfortunately not used or shown for any of the analyses. How is the spread different between LR and HR? Is the reanalysis within the spread?

7) When analyzing absolute numbers (here for blocking, cyclones and windstorms) ties have to be considered in the correlation calculation, ie seasons with the same number of events. Presumably ties are not taken into account as the manuscript does not mention it. 2 possible solutions: i) mask regions with a large number of ties ii) use a different correlation coefficient, e.g. Kendall's Tau B. Otherwise the correlation value could be misleading and statistical significance becomes meaningless, especially in regions with few events per season. There is a significant negative correlation in windstorm frequency in LR over Eastern Canada and a significant positive correlation in HR over the same region. This could be an example of too many ties.

[Printer-friendly version](#)[Discussion paper](#)

8) Related to the above point: Cyclone frequency is apparently masked in regions with high orography. This can be seen in Fig 5. Why is there no mask in Fig 3? What about wind storms. Why are windstorms not masked? Please also consider masking regions with few events per season. There is a mask for blocking. Please state why.

9) The discussion is not critical enough. The reader gets the overall impression of a nearly perfect prediction system regarding the analyzed quantities. Mentioning the correlation value could sometimes already be enough to put the results in perspective. There are some inconsistencies as mentioned above that should be discussed. There is only one sentence P16, L15ff with reference to previous studies with similar objectives. Please add some references or state the lack thereof. See also point 4)

Minor comments:

i) The title suggests an analysis of the entire NH. Please correct. Consider adding the word “deterministic” in the title

ii) “Anomaly correlation” and “skill” are used as synonyms throughout the manuscript. Please state that deterministic skill is assessed through anomaly correlation somewhere in the paper and in the abstract.

iii) There is no reference for the “common shortcoming of climate models” of a too zonal stormtrack in the introduction.

iv) P1L1: The acronym MiKlip is not explained

v) P1L8: “functional chain” is not clear

vi) P1L11ff: Newfoundland is not “downstream” of the stormtrack.

vii) P1L20ff: Please add reference for this paragraph

viii) P2L8: “sectors” is most likely the wrong word

ix) P2L20ff: restructure sentence: “One result...”

Printer-friendly version

Discussion paper



- x) P3L1: specify “lower resolution”
- xi) P3L8: “functional chain?”
- xii) P3L22: change “variables” to “diagnostics” or similar. Variable is not the correct word.
- xiii) P3L29: same, please check the entire manuscript
- xiv) P4L12ff: “However...”: Please rephrase
- xv) P5L1: add “deterministic”. See points i) and ii)
- xvi) P3L2: centered or uncentered anomaly correlation? See point 1)
- xvii) P6L32ff: This is unclear and probably wrong somehow. What kind of percentile is used? Is it the same one in LR and HR? This might explain why the difference in cyclone frequency is not apparent in windstorms
- xviii) P7L2ff: change “nicely illustrated”
- xix) P7L31: the value in brackets is easily misunderstood. Maybe: -3% of a total of X% days in one season
- xx) P10L18: change “implying” to “could be due to” or similar
- xxi) P12L4ff: see point 2 for the whole paragraph
- xxii) P12L22ff: see point 2
- xxiii) P13L35ff: improvement in cyclone frequency improves windstorm frequency? Specifically along the European western coast? P10L12ff highlights the differences of the 2 diagnostics
- xxiv) P15L8: Muller et al 2018 show a decrease of MSLP bias in the Eastern North Atlantic but an increase in the Western North Atlantic in HR. It is therefore only partially “in line”.

[Printer-friendly version](#)[Discussion paper](#)

xxv) P15L25ff: see point 2, for blocking + cyclones

xxvi) P5L10: Please provide a reference or calculation method for the statistical significance. Is the calculation method different between correlation significance and significance for the differences in correlation?

Interactive comment on Earth Syst. Dynam. Discuss., <https://doi.org/10.5194/esd-2019-18>, 2019.

Printer-friendly version

Discussion paper

