

## ***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.***

**Mareike Schuster et al.**

mareike.schuster@met.fu-berlin.de

Received and published: 10 July 2019

Response to Comments of Anonymous Referee 1 [R1]

- R1 – comment 1:  
p1, l.20: the first sentences sound as if extra-tropical circulation is important because it may be linked to extreme events. Isn't it important in a more general sense? After all, it is not a paper on extremes. Will be good to discuss the motivation in a broader context

C1

### **Response to R1 – comment 1:**

The paper is partly on extremes, as windstorms are identified by the exceedance of the local 98th percentile of the surface wind and blocking is identified by blocked flow for min. 4 consecutive days – these are extreme events by definition. However, the extratropical circulation is indeed important in a more general sense - thank you for the remark. The original text was changed to: “The extra-tropical circulation plays an important role for the redistribution of energy in the atmosphere. The prevailing westerlies and the embedded cyclones and anticyclones determine the weather and climate of the mid-latitudes, assisting in balancing temperature and humidity contrasts between the tropics and polar regions. Natural climate variability as well as externally forced climate change determine fluctuations in the circulation and thus i.a. the frequency of extremes such as strong cyclones, intense windstorms or phases of blocked flow. The consequences of such features include extremes in temperature, precipitation/drought and wind speed, often accompanied by immense damage and harm (e.g. Leckebusch2004, Ulbrich2009, Sillmann2009, Pfahl2012, DeutscheRueck2018).”

- R1 – comment 2:  
The term ‘stormtrack’ is confusing when used along with the cyclone frequencies – they are sometimes used interchangeably (not in this paper). Though the Methods describe what is meant by the stormtrack, I recommend commenting on the difference early in the manuscript (maybe even in the abstract)

### **Response to R1 – comment 2:**

p.1, l.3 was changed to: “Four metrics - the Eulerian stormtrack, and Lagrangian blocking, cyclone and windstorm frequencies - are analyzed ...”. We believe that this phrase makes it clear that we differ between the two metrics. The reader can learn about the details of the methodologies in section 2.2.

- R1 – comment 3:  
The same goes to lead years/winters - it is worthwhile explaining which months

C2

are considered. I only found this information in Figure captions

**Response to R1 – comment 3:**

The fact that we are analyzing the winter half year (Oct-Mar) is e.g. stated at p.4, l.6: "We will therefore focus on the winter circulation and evaluate averages of the stormtrack and blocking, cyclone and windstorm frequencies from October through March."

However, to comply with both reviewers' requests for more specific information on the evaluation procedure, the entire paragraph was revised to be more precise and now reads (p.5, l.1ff):

"To derive the deterministic skill of the two forecast systems, we focus on the temporal variability and analyze the anomaly correlation for the winters 2-5 (Oct-Mar), following the Decadal Climate Prediction Project (DCPP, Boer2016) protocol. That means that we calculate lead time dependent anomalies of the circulation measures. This is a simple and robust approach to account for a possible lead time dependent mean bias, i.e. drift. For each of the initialization experiments (1978, 1979, ...) the ensemble average (5 members) of the temporal mean of the 4 contained lead winters is calculated per grid point. This forms a new ensemble mean time series of the lead winters 2-5. This time series serves to calculate the climatology (temporal mean) and to calculate the respective anomaly time series. The time series of those anomalies of the hindcasts is then correlated (Pearson) to the time series of anomalies of the reanalysis. In decadal prediction studies, this procedure is usually repeated for each lead time, thus lead year 1, lead year 2-5, lead year 6-9 - it is therefore referred to as lead time dependent anomaly correlation. In our study we only show results for one lead time: lead winters 2-5. The initialization of the hindcasts takes place in October, this means the first full winter that we analyze is the second winter, i.e. the months 12-17 (Oct-Mar) after initialization. This evaluation procedure is part of the decadal climate prediction evaluation software that was designed within the MiKlip project (Illing2014) and is applied for this study."

C3

- R1 – comment 4:

p2, l28: comment on what parametric bias adjustment approach is.

**Response to R1 – comment 4:**

The wording parametric or non-parametric corresponds to the way how to adjust lead time dependent bias (drift). On the one hand, it is feasible to assume a lead time dependent bias and to fit a curve. Kruschke et al. (2016) called this approach parametric. On the other hand, DCPD recommends to calculate lead time dependent anomalies for each lead year separately (Boer et al., 2016). This is a non-parametric approach. We used the DCPD recommendation in our manuscript.

- R1 – comment 5:

p10,l1: The word 'shift' often implies change in time, consider revising

**Response to R1 – comment 5:**

We think that the structure of the sentence makes it clear, that a spatial shift in LR compared to the reanalysis is meant.

- R1 – comment 6:

p10,l13: I would be more precise here and stick to the words used in the Methodology, i.e. 'open' and 'closed'. Otherwise, you need to clarify what you mean by weak/strong cyclones.

**Response to R1 – comment 6:**

The following sentence was added to p.6, l.20: "Only cyclones that lived for more than 24 hours and reached a Laplace larger than  $0.7\text{hPa}/(\text{degree latitude})^2$  at least once during their lifetime are selected for evaluation."

p.10 l.13 was changed to: "However, it should be highlighted that the cyclone tracking algorithm also detects cyclones in their weak phase, as long as they reach the  $0.7\text{hPa}/(\text{degree latitude})^2$  threshold once during their lifetime." This means, that a cyclone can be part of the evaluation, which lives a couple of days but is generally weak in terms of its Laplacian of the pressure ( $< 0.7\text{hPa}/(\text{degree}$

C4

latitude)<sup>2</sup>), but it reached the intensity criterion for exactly one time step and therefore was included in the evaluation.

- R1 – comment 7:

p.10,l14: I would like to see a figure confirming that positive bias is due to the weak and/or short leaved cyclones. p.10,l.15-16: How do you explain then negative windstorm vs positive stormtrack anomaly over the Atlantic?

**Response to R1 – comment 7:**

To answer this question we selected and evaluated cyclones that, at any time during their lifetime, pass through the central North Atlantic (50-10°W,40°-60°N) - the region where the bias in Fig. 3a is strongest. This analysis is performed for individual cyclone tracks of all initialization experiments between 1960-2012, all 9 forecast winters and all 5 members, for LR and HR respectively.

The intensity histogram (review response Fig. 1) of cyclones that pass the central North Atlantic shows that weak cyclones are more numerous in LR than in HR or ERA-Interim. Although LR overestimates the frequency of weak cyclones in that region, the frequency of the strong cyclones, in that case the strongest 5% of cyclones, i.e. bars to the right of the dashed line, is reproduced quite well in LR. This threshold (dashed lines) in HR (2.87 hPa/(deg.lat.)<sup>2</sup>) is closer to ERA-Interim (2.98 hPa/(deg.lat.)<sup>2</sup>) than LR (2.65 hPa/(deg.lat.)<sup>2</sup>) is to ERA-Interim - but mainly due to the generally larger number of events in LR. Overall, the shape of the intensity distribution for cyclones passing that region is much more similar, and actually almost identical, between HR and ERA-Interim, than between LR and ERA-Interim. The lifetime histogram (review response Fig. 2) of cyclones that pass the North Atlantic region also shows that short-lived cyclones in this region are more frequent in LR than in HR or ERA-Interim. Again the shape of the distribution matches very well between ERA-Interim and HR.

Regarding the second part of the comment: The stormtrack is calculated from the variance of the geopotential height in the synoptic band. Within this quantity,

C5

there are many systems included which do not produce windstorms, as this is the variability of all geopotential values (high pressure as well as low pressure systems, and strong ones as well as weak ones). Only strong low pressure systems can be related to windstorms. It cannot be expected that the signals of total variance of geopotential height (stormtrack) are identical to those of the windstorms since the distribution of strong and weak cyclones changes differently as discussed above.

- R1 – comment 8:

p10, l.31: I can see a discussion on negative correlations further in this section (e.g.p13, l7)

**Response to R1 – comment 8:**

As we lined out in the paper, it is not desirable to have a deterministic prediction model to continuously predict the opposite of the observed quantity. We therefore stick to the opinion that negative correlations should not be considered skillful and will therefore not discuss them in detail. If anything, then the message of our paper is that the amount of negative correlations is reduced in HR. This is covered by the discussion of positive differences between HR and LR (Fig. 4e, 4f, 5e, 5f).

- R1 – comment 9:

p12,l6: I How about a strong reduction of skill over Northern Canada and the Barents Sea

**Response to R1 – comment 9:**

The following sentence was added to p.12, l.6: “However, there is also an area of significant reduction of the anomaly correlation for the stormtrack over Northern Canada and the Baffin Bay.”

- R1 – comment 10:

p.12, l3: 'significant skill improvement' - the authors probably mean that HR model

C6

shows statistically significant correlation with ERA-Interim at some points. In my opinion, though, this statement makes an impression that skills of model prediction have become really good (so say at least 'statistically significant skill improvement' or rephrase). More important, the prediction skills, as shown in the paper, are remarkably low for most part of the region, but this message is not conveyed by the paper - will be good to see more discussion on that.

**Response to R1 – comment 10:**

Referring to Fig. 4.e in this line (p.12 l.3), we are not discussing the skill of HR compared to the reanalysis but rather the change in skill from LR to HR. Thus, we indeed mean that the change from LR to HR shows an improvement, e.g. from low or (significant) negative correlations (no skill) in LR to significant positive correlations (skill) in HR, i.e. an improvement in skill or one could also say a statistically significant improvement in anomaly correlation.

Please note that deterministic decadal prediction skill in terms of anomaly correlation is generally low for model variables other than surface temperature (compare skill of precipitation in Kadow et al., 2016 or cyclone frequency in Kadow et al. 2017 to skill of surface temperature in Pohlmann et al., 2013).

We changed the original wording to: "...statistically significant skill improvement..." and added values of correlation in the text to put results into perspective.

- R1 – comment 11:

Figures 4-5: In line with the previous comment, it will be interesting to calculate the percentage of area that is significantly (positively?) correlated with ERA-Interim. This number can be added to each subplot.

**Response to R1 – comment 11:**

As the climatologies of the circulation quantities show, the values of the anomaly-correlation are not equally important anywhere in the displayed domain. A significant positive correlation e.g. along the maximum blocking frequency is more relevant than a significant positive correlation over a region with very low frequen-

C7

cies. Therefore, we think the suggested percentage of grid points with significant positive correlation would be misleading, as it would weigh "irrelevant" regions equally to "relevant" regions. We produced the figure (review response Fig. 3) to answer the reviewer's question but will not show it in the paper for the stated reasons. The figure shows that for all circulation quantities the number of grid points (in the North Atlantic domain) with significant negative anomaly correlation is reduced in the higher resolution system and the number of grid points with significant positive anomaly correlation is increased in the higher resolution system - supporting the theory of improved physical processes throughout the region.

We understand the referee's main point with this comment is, similar to R1 comment 10, that we emphasize the positive effects of the resolution and the reader could think we suggest that HR is the perfect decadal prediction model. This is however not our intention. We rephrased respective sentences.

- R1 – comment 12:

Discussion and Conclusions: this section is too long, consider shortening. Parts of the discussion may be moved to the Results. The last sentence of the article is not clear, please revise.

**Response to R1 – comment 12:**

Last sentence was changed to: "An important question, that should be addressed in future studies is: Which physical processes form the foundation of this detected decadal prediction skill for the different circulation variables?"

- R1 – comment 13:

Fig4: significant at what level

**Response to R1 – comment 13:**

We added "(95% significance level)" to p.5 l.10.

- R1 – comment 14:

p2,l.11: remove comma before dash. p2, l18-19: put references in brackets p2,

C8

l29: did you mean more skillfull ? Skillful is misspelled. p5, l11: 1000 time - should '1000 time steps' be better? p10, l6: should read 'these results'

**Response to R1 – comment 14:**

We replaced the dash with a comma, to separate the two independent but related sentences.

Thank you for noticing, brackets were inserted.

No, we do mean simply skillful.

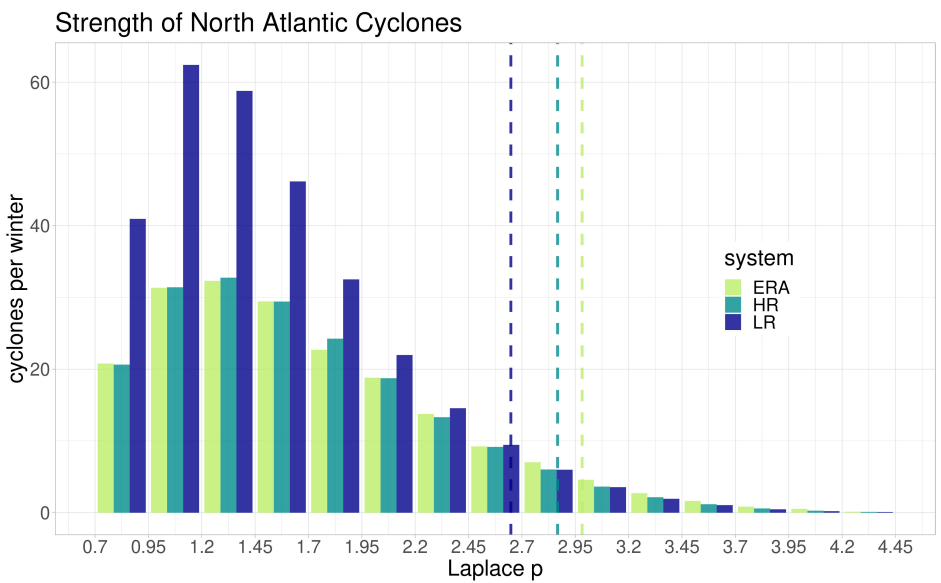
We are using American English throughout the paper, the spelling is correct: skillful.

We use the bootstrap method, to estimate the distribution of a population by resampling the dataset with replacement. This is repeated 1000 times. The suggested term “1000 time steps” is not applicable.

The word “results” is used as a verb here. The suggested use of the word “results” as a noun would change the meaning of the sentence and leave it incomplete.

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

C9



**Fig. 1.**

C10

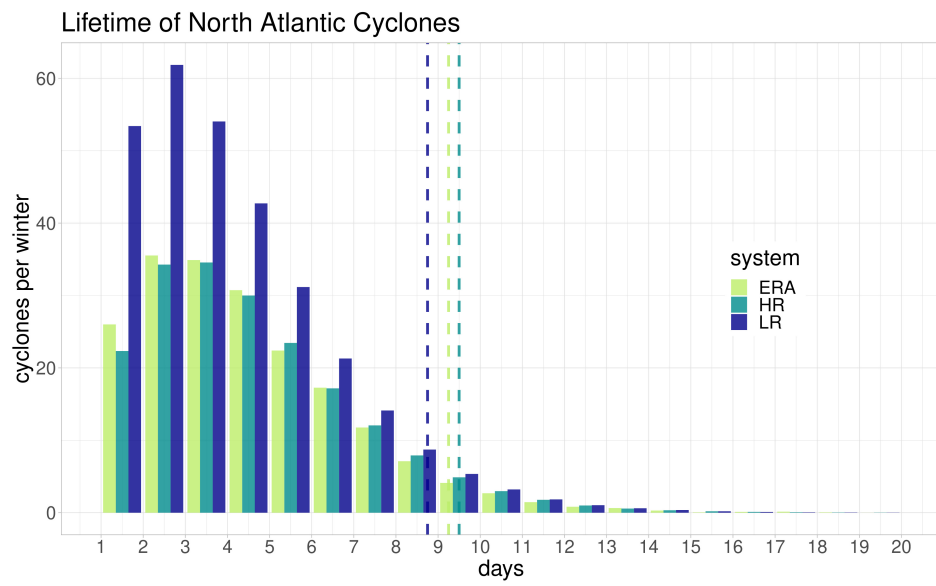


Fig. 2.

C11

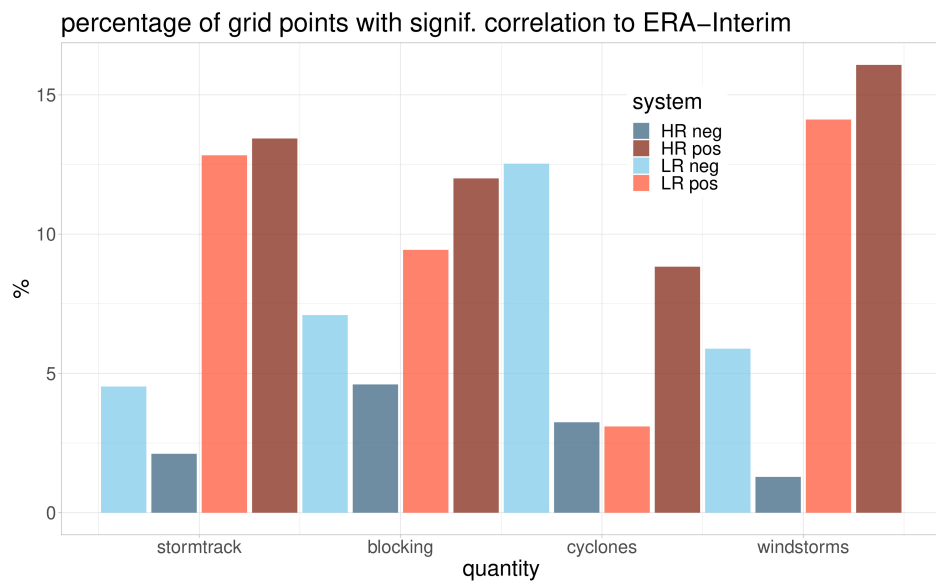


Fig. 3.

C12