

**Review of “Using a nested single-model large ensemble to assess the internal variability of the North Atlantic Oscillation and its climate implications for Central Europe”
by A. Bohnisch et al.**

The manuscript presents an analysis of changes in the North Atlantic Oscillation (NAO) under a global warming scenario, using two 50-member model ensembles: an ensemble of a global general circulation model, and an ensemble of a high-resolution nested regional climate model. The large ensemble size allows the authors to not only analyze the change in the mean NAO, but also in its variability. The authors also show the impact of the NAO and its variability on European climate.

This manuscript presents an interesting study that combines two state-of-the-art techniques: very large ensembles to estimate transient change of internal climate variability, and a high-resolution regional climate model. The results are novel and relevant. However, I think there is some unused potential in the study that should be harvested (see my specific comments below), and the presentation of the results could be improved.

I think the manuscript is a good fit for Earth System Dynamics and should be published. That being said, the manuscript requires structural clarification that warrants a major rewrite, so that I recommend **major revisions** to the manuscript before publication can be considered.

Specific Comments:

I. 2 “...(NAO) which is a relevant index for quantifying natural variability...” I find this sentence to be ambiguous. What is a relevant index? As it stands now, it seems to be the mass advection triggered by the NAO. I suspect that the authors mean the NAO itself. If this is the case, I think this ambiguity can be avoided by introducing a comma between “(NAO)” and “which”.

I. 4 Is the link to the CORDEX project really needed in the abstract? Please consider removing it.

II. 4-6 This sentence is missing the crucial information that the “LE” model is a nested regional climate model.

I. 9 I do not see how the word “strength” in brackets on its own relates to “pearson correlation coefficient”. Please re-evaluate whether “strength” adds any meaning at this point.

I. 11 What is a “correct response” to NAO forcing? How is that defined? If it’s based on the global model simulation (which I assume it is) I am not sure that “correct” is the right word here.

I. 12 Which relationships weaken in the future? Also, what does it mean and why is it important to show that the amplitude of inter-member spread does not change with anthropogenic forcing?

Introduction I find the introduction confusing and hard to follow. For example, the first paragraph (II. 16-22) seems to set the reader up for a following paragraph on ensembles, but instead global and regional climate and the NAO are introduced in the next paragraph (II. 23-32). For another example, the reader expects a discussion of advantages and limitations of different methods to quantify the NAO index after paragraph 3 (II. 33-37), but paragraph 4 (II. 38-42) introduces the reader to NAO impacts and its interactions with other modes of climate variability. Moreover, this interaction with other modes of variability is in my opinion not important to the study presented in this manuscript. Both the missing storyline and the lack of focus on the important information for this study are an issue throughout the entire introduction. I therefore recommend that the authors rewrite the introduction with particular attention to the storyline and focusing on the important information, so that the reader can follow the reasoning more easily.

I. 38 There is no mention of a positive state before. I believe the authors are referring to a positive NAO state, but that needs to be made explicit, especially so at the very beginning of a paragraph.

II. 75-76 Please consider omitting the “table of contents” at the end of the introduction. It does not add to the story and takes focus off the nice overview of key questions that will be addressed in the paper just before.

II. 80-86 I think somewhere here it would be important to mention which region the regional model covers. Please consider adding this crucial information.

I. 91 The implications of this sentence would be much easier to understand, if the CORDEX ensemble was introduced very briefly. Please consider adding a few words on what the CORDEX ensemble is, as well as a literature reference.

II. 95-96 I am not sure that I agree with the conclusion, that “the most important” modes of climate variability are captured by the ClimEx model, as this conclusion is here based on a comparison to another model ensemble. I agree that it is reasonable to assume from this comparison that the ClimEx model produces reasonable climate variability, but I do not think such a comparison warrants a judgment on which mode of variability is important or not. Please consider rephrasing.

II. 100-103 The most commonly used acronyms for sea level pressure and surface air temperature are SLP and SAT, respectively. Why did the authors decide to use different abbreviations? This is not a huge issue, but interrupts the flow when reading. Also, t2m and tas are usually not the same in model output. The manuscript would benefit from clarification as to which of the two is used in this study – this is currently not clear.

I. 120 The text says that there are two regions of interest, while table 2 specifies seven regions and the remaining manuscript references those seven regions. I suggest omitting the “two regions” phrase, as it is more confusing than helpful at this point.

I. 140 The authors use past tense to describe the present study here, and this appears to be the dominant choice of tense. Elsewhere, however, present tense is used (e.g. I. 120 “...there are two separated regions...”). This inconsistency can be found throughout the entire manuscript. To improve readability, I suggest the authors decide on one tense and stick to it throughout the manuscript.

II. 141-142 The word “representative” is lacking a reference here. The 30-year time horizon leads to an NAO distribution that is representative of what? Please elaborate briefly.

II. 144-145 This is an important caveat. I like that this is mentioned here, but missed it in the discussion section. I suggest taking it up again there to make sure this (perfectly acceptable) limitation of the study can be appreciated.

II. 150-154 I think this bit would be easier to understand if the order of the phrases was altered to first explain why March can be included and then say that DJFM is used for winter. Please consider making this change.

II. 159 I suggest refraining from the statement that a station-based NAO index is “easy” to interpret – its reference is arbitrary (easy for whom?) and it is not a very scientific expression. Please rephrase.

II. 189-195 This section appears to already present results. Please consider moving it to the results section.

I.200 In lines 97-98, the authors define REF as the ERA-Interim data set. Here, REF appears to refer to the NAO index within the ERA-I data set. Please define REF only once and unambiguously.

II. 205-206 I am not sure I agree that figure 1a shows that REF (the blue bars) lies “comfortably” within the ensemble spread (grey & red). Particularly negative extremes, but to some degree also positive ones, seem to be underrepresented in the model. Can you please comment on this and possible implications for this study?

I. 214 “...original data into three subsets...”

I. 214 Please consider changing “indifferent” to “neutral” or “average” here and throughout the document.

I. 214 Are the “average psl conditions” referenced here the same as the “MSLP mean” in figure 2? If so, I highly recommend using coherent names (i.e. “mean” or “average” in both cases) to avoid confusion. I had to read this paragraph several times before I understood it.

- II. 216-217 Which difference is referenced here? Also, what do over- and underestimation refer to? If this is based on a comparison of figs. 2a and d, I cannot follow the argumentation – actually, it appears to me that the model overestimates mean SLP over the North Sea and underestimates SLP over Greenland. Can you please clarify?
- I. 218 “...phases also show less pronounced...” Weren't the anomalies more pronounced in the model than in REF for the mean state? If so, please omit the “also”.
- II. 239-240 “...the spatial patterns of ERA-I and CRCM5/ERA-I differ **more strongly than in Fig. 3...**”
- I. 241 What is the reference for the “more humid conditions”? The lack of a reference for relative statements is an issue that needs addressing throughout the manuscript.
- I. 256 The NAO explains less variance than what?
- I. 257 tas std decreases less than what?
- I. 259 While I am sure the inconsiderable change of spatial patterns compares the historical to the projected period, I think it would help to give this information here again.
- II. 259-260 Could you please give a figure reference for the claims made here?
- I. 264 Is there a particular area for which the transfer of internal variability from GCM to RCM is assessed?
- I. 277 If large tas deviations do not correspond to high or low α , what do they correspond to?
- I. 284 I find the presentation of this reference to figs. 3, 4 and 5, h & i ambiguous. Do you refer to panels h & i of all those plots, or just 5?
- II. 301-302 This sentence is difficult to understand due to the many parentheses and different references therein. I highly recommend splitting this sentence in at least two.
- I. 306 I think the “matching subset region time series” warrant a more detailed explanation. As it stands, I am not sure what these are and how to interpret them. As a result I cannot follow the text. Please introduce this metric at least shortly.
- II. 308-309 I am not sure I fully agree with this statement. While correlations indeed appear to be generally lower for pr sum (fig. 8b), 1/3 regions for tas mean (fig. 8a) and 2/3 regions for tas std (fig. 8c) show an increase towards the later period. I think the manuscript could benefit from a more detailed discussion here.
- II. 309-310 I do not quite understand the last sentence of the “results” section. As a result, I struggle to see what its consequences are. I recommend adding some more explanation here, as this might be a crucial point.
- II. 314-315 What does it tell us that one realization shows a good correlation to REF? Why are the two so highly correlated? I am not sure why this is mentioned here. As in the introduction, this (apparently) irrelevant information might cause the reader to lose track of what is important. Please consider omitting this sentence or, if you deem it relevant enough, elaborate to illustrate its relevance.
- I. 316 It is not clear about which strong psl gradient the authors are writing here.
- II. 318-319 NAO+ and NAO- are weaker within CanESM2-LE than which reference?
- I. 320 The very limited sample size of n=7 (or rather n=3 and n=4) in REF is an important issue that is worrisome. It should be discussed further! How robust are the results presented here? What could maybe be learned about observations from the model?

I. 326 At this point, I somewhat expected a discussion on the influence of other teleconnection patterns. I think the authors should at least provide some indication (from the literature) about how large these teleconnections' influence on this study can be expected to be.

I. 335 The latter is not as clear in the chose domain as what?

II. 338-339 I think the observation is missing a reference in this sentence: Is it NAO+ or NAO-? And are these observations derived from reanalysis or the literature or a model? As it stands, this is quite ambiguous.

I. 350 Omit the comma between "region" and "which".

II. 352-353 This is an intriguing thought. What are its consequences/implications? Please consider to elaborate a bit.

I. 361 What does it mean for the findings presented here that the GCM overestimates T and pr? Does this limit the conclusions that can be drawn?

I. 367 Since the patterns are "only" very similar, I find the statement "atmospheric dynamics are correctly implemented" a bit too strong. Please consider rephrasing to, e.g., "...can be regarded as correctly implemented".

I. 378 As stated before (comment lines 205-206), I do not agree that the observations lie comfortably within the model spread, so I also have an issue with the statement "...the same climate statistics". Please either explain where I went wrong or rephrase.

II. 382-383 Maybe rephrase to "...with highest change in CRCM5-LE, but not **necessarily** in CasESM2-LE."?

I. 391 Less tas and pr variation is explained by NAO than by what?

Conclusions I think the reference to the questions raised in the introduction could be made clearer. While the references are there, I think it would make this part clearer if it was structured in bullet points, like the questions raised in the introduction. Please consider making this change.

II. 397-399 This is a long sentence that is hard to understand because it takes up two different points. Please consider splitting the sentence in two.

I. 404 I find the word "proves" very strong. I agree that the clearly visible topographic features are nice to look at and encouraging for the model presented here, but I disagree with the notion that the mere notice of more pronounced topographic features "proves" the added value of anything. High resolution does not always equal added value. Please rephrase.

Fig. 2 caption "(g)-(j): 2070-2099 changes **with respect to 1981-2010**"

Figs. 3-5 caption What are the correlations show in blue isolines? What is correlated to what? Also, this is a confusing figure, partly due to the ambiguous headers for the subpanels (which are identical for, e.g., c and g). Please think about a more intuitive way to convey this very interesting information.

Fig. 6 Some of the indices named in the upper left corner have slightly different names than those found on the x-axis. It could help the clarity of the (otherwise very nice and interesting!) figure if those names were the same. Please consider changing the figure accordingly.

Fig. 8 Please explain a, b and c in the caption. Also, I do not quite understand what is displayed. What is a "similarity of matching regions"?

Fig. A2 caption Please explain the subpanels in the caption.