

Detailed Responses to Referee Comment No. 1

Dear Reviewer No. 1,

We would like to thank you very much for your constructive and valuable comments. We think that they will considerably improve the scientific quality of our manuscript. Below, we answer to all points raised in the Referee Comment (comments in grey, answers in black). Please consider also the responses to Reviewer No. 2 as there are some cross-overs regarding the comments.

With kind regards,
Andrea Böhnisch on behalf of all co-authors
12 December 2019

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.

Thanks for the generally positive reception of the manuscript. We agree that the manuscript needs further structural improvement which was mentioned by both reviewers. We will work on the presentation of the results in order to harvest the unused potential; please see responses to the specific comments below. We also intend to include an analysis of the large-scale SLP pattern in the RCM data (see Major Comment of Reviewer No. 2) as we think that this may help us to interpret NAO responses in central Europe.

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”.

Yes, the second part of this sentence is referring to the NAO. We will introduce a comma to clarify.

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

As the link to the ClimEx project is also included in the Data section, it is not needed in the abstract. We will remove it.

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

This is true. The missing information will be included.

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.

The word “strength” refers to “strength of linear relationship”. As it does not add any meaning in the abstract, we may remove it here.

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.

We agree that “correct” is ambiguous in this context. To underline the intended meaning, we will rephrase to “... responses which are comparable with reference reanalysis data”.

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?

This sentence refers to the relationships between the NAO and corresponding responses. The finding that the amplitude of the inter-member spread does not change suggests that internal variability of responses and uncertainty of response assessment are similar in both time periods. – We will include the reference and add a corresponding explanation.

Introduction I find the introduction confusing and hard to follow. For example, the first paragraph (Il. 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 (Il. 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 (Il. 33-37), but paragraph 4 (Il. 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.

We agree that the introduction is not written as clearly as it should be. So we will restructure the section in the attempt to focus on the four major topics of interest – internal variability, the NAO, nesting and ensemble approaches. We will open our introduction with the explanation of internal variability, introduce the NAO as a mode of internal climate variability, continue with the NAO representation in various climate models, introduce the ensemble approach with which one can assess potential NAO shapes (i.e. the model internal variability of the NAO) and close with the necessity of regional climate models when analyzing NAO responses in heterogeneous regions. We will also attempt to clarify the study goals within the four topics. The key questions will integrate the four topics and represent the structure of the following analyses.

We will also remove the information on interactions of the NAO with other modes of climate variability, as it indeed is not crucial for the study.

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.

Yes, that is true. The “positive state” refers to the positive NAO state introduced in lines 26-27. To improve the readability, we will change the beginning of this paragraph to: “Compared to the neutral state, the positive NAO state leads to warmer ...”

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.

It is true that the “table of content” does not add important information. We will remove it.

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.

Currently, the information on the region follows in lines 87-90. We will change the order of some phrases to provide the information on cover regions earlier.

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.

Thanks, this is a good hint. Other than the ClimEx ensemble, the CORDEX ensemble consists of several GCM-RCM combinations set up in a coordinated modelling framework, and aims at evaluating model variability. We will include a short summary on the CORDEX ensemble, based e.g. on Giorgi et al. 2009.

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.

We agree that the focus of this paragraph should not be set on the judgement of importance of modes. We will rephrase the sentence accordingly.

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.

Thanks for this hint. We will change the names from psl → SLP, tas → nSAT (near surface air temperature) and pr → PR whenever the variable is meant. Table 1 introduces the model output variable names, which is why we will keep psl, tas, pr etc. in there.

In CanESM2 and CRCM5 “tas” refers to near surface air temperature, and ERA-I variable “t2m” is 2-m temperature. We assumed that t2m is the ERA-I variable that is most similar to the model variable. We will place an explanation in the manuscript.

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.

Originally, the “two regions” in this phrase refer to the NAO formation (1) and response (2) regions. We agree that the mention of seven analysis regions in Table 2 does not fit the “two regions” phrase, so we will remove it. Following a suggestion of Referee Comment 2, we will also replace Table 2 with a labeled map indicating the size and position of the regions of interest.

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.

Thanks, we will fix this.

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.

As stated in the sentence before (lines 140-141), major fluctuations of the natural climate system on several temporal scales are assumed to be included in the 30-year time horizon. Their potential influence on the NAO may thus be seen as represented within the sampled NAO time series. – We will rephrase these sentences (lines 140-142) in order to clarify.

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.

Thanks, we will include this caveat in the discussion section.

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.

Thanks, we will change the order of the sentences.

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.

We agree that “easy to interpret” is not an appropriate expression in this context. We will rephrase the paragraph to justify our choice of calculation method.

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

This paragraph was included to explicitly mention the way internal variability was addressed in this study. It is not intended to present results. We will change the phrasing accordingly to clarify.

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.

Yes, reference/REF is defined to be anything derived from the ERA-I data set, but this sentence uses REF confusingly. Therefore, we will change the sentence to “... a reference NAO index was derived from ERA-I ...”

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? *This sentence refers primarily to the x-axis of the histogram in Fig. 1a, not the frequency of occurrences: The index values of the ERA-I NAO index may be found within the CanESM2-LE, that is, between the minimum and maximum index values. It is true, that the distribution of ERA-I index values shows differences towards the distribution of the CanESM2-LE. These differences may partly be explained by the different sample sizes ($n_{\text{ERA-I}} = 30$, $n_{\text{CanESM2-LE}} = 1500$, we will stress this information in the text); the ERA-I sample is only one realization which is compared with 50 realizations of the ensemble, so deviations between the distributions may occur. A better way to display these results may be the usage of CDFs which is why we will change the type of this figure.*

I. 214 “...original data into three subsets...”
Thanks, we will fix it!

I. 214 Please consider changing “indifferent” to “neutral” or “average” here and throughout the document.
Thanks, we will change it to “neutral”.

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.
Yes, both refer to the same. We will adopt your suggestion and change the wording in both text and Fig. header to “neutral SLP conditions”.

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?
“Difference” refers to the mean SLP difference between CanESM2-LE and ERA-I (Figs. 2a and 2d, respectively). MSLP over Greenland rises to about 1025 hPa in CanESM2-LE and about 1015 hPa in ERA-I data (hence overestimation in CanESM2 with respect to ERA-I); over the North Sea, MSLP reaches 1000 hPa in the CanESM2-LE and 1010 hPa in ERA-I (hence underestimation in CanESM2 with respect to ERA-I). We will clarify the wording and change the coloring of the MSLP maps to better visualize the differences.

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”.
That is true, we will remove the “also”.

II. 239-240 “...the spatial patterns of ERA-I and CRCM5/ERA-I differ **more** strongly than in Fig. 3,...”
Thanks! We will fix this.

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.
In this case, the reference is the neutral NAO state. Where necessary, we will include the respective references to clarify (see also following comments regarding the same issue).

I. 256 The NAO explains less variance than what?
In this case, a temporal comparison was drawn. We will change the sentence to: “...less variance under future conditions compared to the historical period”.

I. 257 tas std decreases less than what?
This is also a comparison between both time horizons. We will include: “...in the projected future climate”.

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.
We will add this information.

- II. 259-260 Could you please give a figure reference for the claims made here?
Of course: the claims refer to panels (g) & (i) in Figs. 3-5. We will include the figure reference in the manuscript.
- I. 264 Is there a particular area for which the transfer of internal variability from GCM to RCM is assessed?
We assessed the “transfer” in the response regions – that is, spatially explicit in CEUR (see std.dev50 maps/subpanels (d), (f) in Figs. 3-5) and spatially aggregated in NE, BY, SE (see Fig. 7). A corresponding note will be placed in the manuscript.
- I. 277 If large tas deviations do not correspond to high or low α , what do they correspond to?
Thank you for this question; this sentence is not as clear and detailed as it should be. We will include the correlations between the α_1 and std.dev50 values to give a more differentiated impression of the relationship between α_1 values and the inter-member spread.
- 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?
Yes, you are right; the reference is ambiguous. Looking in figs. 3-5, we also noted that there is a mistake; it should be (j), not (i). To clarify that we meant panels (h) & (j) of figs. 3-5, we will change the reference to “panels (h), (j) in Figs. 3-5”.
- 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.
Thank you, we will rephrase this sentence and also the preceding sentences that suffer from the same problem.
- 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.
This is a good point. The idea was to compare the variability of tas mean, pr sum and tas std time series of the CRMC5 with the CanESM2 in the subset regions NE, BY, SE. Therefore, we correlated the time series of, e.g. tas mean, derived from the spatially aggregated subset region in CRMC5 with the time series derived from the CanESM2 subset region. These correlations were calculated member-wise, leading to 50 correlation coefficients per subset region. High (low) correlation coefficients indicate a strong (weak) co-variability of the CRMC5 and CanESM2 in the respective member. We will add a brief explanation in the manuscript. Therefore we will rephrase this paragraph (306-310) and also incorporate changes following the next two comments.
- 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.
These findings are certainly true. We will review and correct the statements, include a more detailed discussion and elaborate on the implications of these findings.
- 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.
The last sentence is not as precise as it should be. The results presented in Fig. 8. suggest that there is a larger discrepancy in portraying pr sum and tas std in the RCM with respect to the GCM than for tas mean. We will rewrite the paragraph concerning the “matching subset region time series” (see also comments to lines 306, 308-309) and thereby also elaborate on this finding.
- 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 loose track of what is important. Please consider omitting this sentence or, if you deem it relevant enough, elaborate to illustrate its relevance.
This realization was mentioned to show that the ensemble may incidentally produce very “realistic” looking realizations. However, we agree that it might seem irrelevant and distracting, so we will remove the sentence.

I. 316 It is not clear about which strong psl gradient the authors are writing here.
Yes, this information is missing here. We refer to the SLP gradient over the North Atlantic within the CanESM2 under neutral SLP conditions as seen in Fig. 2 (d). The sentence will be updated with the corresponding information.

II. 318-319 NAO+ and NAO- are weaker within CanESM2-LE than which reference?
The reference (which is indeed missing) is the ERA-I data set. We will include it in the revised manuscript.

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?

We agree that the small sample size is problematic. We will conduct an uncertainty assessment of the samples for positive and negative NAO composites (referring to Fig. 2, panels (b)-(c) and (e)-(f)). For example, we may estimate the standard error of the arithmetic mean on each grid cell for the ERA-I data and compare it with the CanESM2 samples (which are considerably larger). This analysis should show where the anomaly patterns can be seen as robust in both models.

However, it may be difficult to learn about observations from the model. Learning from the model about observations would imply that the model internal variability can be seen as “correct” as the observed internal variability which is not easy to estimate as there is only a single realization of observations.

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.

This is an interesting suggestions, thanks! We will include a short survey on the influence of the East Atlantic Pattern and the Scandinavian Pattern, as we based our NAO index on the SLP gradient over the North Atlantic which occasionally is affected by these teleconnection patterns (see Moore et al. 2013 and Comas-Bru and McDermott 2014).

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

The sentence was intended to say: “ ... is not as clear as the first in the chosen domain”. We will rephrase accordingly.

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.

We agree that this sentence is ambiguous. It is meant to refer to the fact that the Jetstream position it altered during the NAO+/NAO- phases and therefore associated air mass advection is displaced (see e.g. Woollings et al. 2015). However, as we do not further refer to the Jetstream in the text and thus the sentence does not add to the argumentation in the discussion, we will not necessarily keep it in the revised manuscript.

I. 350 Omit the comma between “region” and “which”.

We will fix this.

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

We will include a more detailed discussion on this subject. The GCM reproduces strongest variability in (geographically) other regions than ERA-I, but in the RCM the positions are “correct”; so for example, we may also see added RCM value for regional scale analysis in this.

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?

This information is given as background information. The overestimation of average T and pr does not affect the findings regarding the correlation coefficient and α_1 as these are based on the changes/variability. The results in the study thus represent the differences in the model-specific variability related to the NAO (which show also some kind of bias regarding stronger/weaker variability in the RCM/GCM). We will update the text accordingly.

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”.

Thanks, we will rephrase the statement accordingly.

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.

Thanks for your concerns. We agree that “the same climate statistics” sounds too strong. As shown in Fig. 6, the CanESM2 ensemble generally encompasses the REF realization regarding several statistics, e.g. inter-annual variability or number and mean values of positive/negative phases. We will insert clearer references in this paragraph and also rephrase to: “The ensemble also shows comparable climate statistics with the REF time series and patterns”.

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

Thanks, we will adopt the wording suggestion.

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

“Less” is referring to a comparison between historical and future time frames. We will rephrase accordingly.

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.

Thanks for this idea. We will put the answers to key questions (a)-(d) in bullet points.

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.

Thank you; we will consider this suggestion.

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.

Thanks for this concern. We agree that “proves” sounds rather strong. We will rephrase to “suggests”.

Fig. 2 caption “(g)-(i): 2070-2099 changes **with respect to 1981-2010**”

Thanks, we will fix it.

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.

The blue isolines correspond to lines of equal correlations between the NAO index and the tas mean/tas std/pr sum time series on the grid cells by increments of 0.1. We agree that the bare presentation of blue isolines is rather confusing. We will change the increments to 0.25 (in order to picture less lines) and indicate the correlation strengths by different grey scales (and a legend). We think that figures 3-5 will gain more clarity in doing so. Also, headers and captions will be changed to better indicate what is shown by the single subpanels. For example, headers may be updated with information on the time frame (1981-2010 for first and second row, 2070-2099 for third row).

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.

Thank you. We will correct the names in the text box.

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”?

These figures display the temporal co-variability of the corresponding CanESM2 and CRCM5 members in the three subset regions (NE, BY, SE) for tas mean (a), pr sum (b) and tas std (c). Thus “matching” refers to the same member in the GCM and RCM.

We will include a detailed description of the metric in the text (see also response to comment line 1. 306), and change the caption accordingly.

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

Figure A2 shows the ratio of tas mean α_1 and winter tas std for the data sets employed in the study: (a) CRCM5/ERA-I and (b) ERA-I under historical conditions, and CanESM2-LE ((c)-(d)) and CRCMR5-LE ((e)-(f)) under historical and future conditions. – We will add a description of the subpanels in the caption.

References in this response:

Comas-Bru, L., McDermott, F. (2014): Impacts of the EA and SCA patterns on the European twentieth century NAO-winter climate relationship. *Quarterly Journal of the Royal Meteorological Society*, 140, 354-363.

Giorgi, F., Jones, C. Asrar, G. R. (2009): Addressing climate information needs at the regional level: the CORDEX framework. *WMO Bulletin*, 58 (3), 175-183.

Moore, G., Renfrew, I., Pickart, R. (2013): Multidecadal Mobility of the North Atlantic Oscillation. *Journal of Climate*, 26, 2453-2466.

Woollings, T., Franzke, C., Hodson, D., Dong, B., Barnes, E., Raible, C., and Pinto, J. (2015): Contrasting interannual and multidecadal NAO variability, *Climate Dynamics*, 45.