Dear reviewers, dear editor,

We would like to thank you for your constructive advice and suggestions. We are confident that these truly improved the quality of our manuscript.
Please find our point-by-point responses to the issues raised in the Referee Comments below. The structure of this document is as follows:

1) Responses to Referee Comment No. 1

2) Responses to Referee Comment No. 2

3) Responses to Editor Comment

4) Revised manuscript with track changes

There are some general remarks that do not apply to any specific comment, but arose during the revision of the manuscript:

- Appendix Figures: We removed (old) Figs. A3, A4 and A6. The information included in Figs. A3 and A4 was inserted in new Figs. 6 and 8 (i.e. spatial distribution of α1; stippling now indicates regions with SNR < 1). Fig. A6 (i.e. 50 GCM index time series for 1981-2010) was not regarded as adding meaningful information to the discussion in the manuscript.

- Additional Figures: In order to clarify that GCM members may be regarded as independent in the early time frame, we included a figure showing the inter-member standard deviations among five-member groups with the same ocean initial conditions and among five-member groups with mixed ocean initial conditions. We also removed Table 2 (position and size of study regions) and present the information in a map.

- Additional abbreviation: special attention was paid to the use of words like “inter-member spread”, “internal variability”, “noise”, “std.dev50” which were used somewhat interchangeably and imprecise in the discussion paper. In order to clarify, we tried to use “inter-member spread” whenever we meant the range (maximum to minimum) of members in the ensemble (new abbreviation: IMS). When referring to the spatially distributed IMS expressed as the standard deviation among the 50 members, we used “ensemble sd” as it does not mean exactly the same as IMS. We also decided to drop the term “std.dev50” which was not used consistently in the manuscript.

- Additional analyses: following a major comment of reviewer No. 2, we included an analysis on the large-scale RCM SLP pattern within the CEUR domain.

- Information on the lines with changes and figure/table references refer to the updated manuscript if not stated otherwise.

- Correspondence: we changed the e-mail address from a.boehnisch@iggf.geo.uni-muenchen.de to a.boehnisch@lmu.de

The comments raised by the referees are marked in blue, responses in black (explanations in italics).

With kind regards,

Andrea Böhnisch on behalf of all co-authors
17 February 2020

Responses to Referee Comment No. 1

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 worked on the presentation and text structure in order to increase readability. For example, the introduction and discussion sections now clearly follow the key questions raised at the end of the introduction. Additional analyses regarding the large-scale SLP pattern present in the RCM data within the European domain were performed, and we aimed at better assessing the uncertainty within the analyses.

Specific Comments:

l. 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".
We included the comma in order to avoid the mentioned ambiguity.

l. 4 Is the link to the CORDEX project really needed in the abstract? Please consider removing it.
In the submitted discussion manuscript, the link to the ClimEx project was also included in the data section, such that it was indeed not needed in the abstract. When rewriting the introduction, we decided to move it to the first mention of the project. The text now reads:

l. 61-63: "Such downscaling of a GCM single-model large ensemble was performed within the Climate Change and Hydrological Extremes project (ClimEx, www.climex-project.org, Leduc et al., 2019)."

ll. 4-6 This sentence is missing the crucial information that the "LE" model is a nested regional climate model.
We updated the sentence accordingly. It now reads:

l. 4-5: "In this study, 50 members of a single-model initial condition large ensemble (LE) of a nested regional climate model were analyzed for a NAO-climate relationship."

l.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.
"Strength" refers to the strength of the linear relationship expressed by the Pearson correlation coefficient, but it does indeed not add additional information. We rewrote this part of the abstract, thereby removing the word "strength":

l. 7-10: "Responses of mean surface air temperature and total precipitation to changes in the index value are expressed for a Central European domain in both the CanESM2-LE and CRCM5-LE via Pearson correlation coefficients and the change per unit index change for historical (1981–2010) and future (2070–2099) winters."

l. 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 rephrased:

l. 11-12: “Reproductions of the NAO flow patterns in the CanESM2-LE trigger responses in the high-resolution CRCM5-LE that are comparable with reference reanalysis data.”

l. 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 included the reference and added a corresponding explanation.

l. 12-13: “NAO–response relationships weaken in the future period, but their inter-member spread shows no significant change.”

l. 488-490: “When comparing present and future values, a vertical shift of the boxes in Fig. 10 indicates that r is reduced in the future, but the inter-quartile distance of the r distributions (box size) stays nearly the same for GCM and RCM. This shows that the uncertainty range of the signals does not change significantly in the future horizon.”

Introduction I find the introduction confusing and hard to follow. For example, the first paragraph (ll. 16-22) seems to set the reader up for a follwing paragraph on ensembles, but instead global and regional climate and the NAO are introduced in the next paragraph (ll. 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 (ll. 33-37), but paragraph 4 (ll. 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 introducion. 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 restructured the introduction in an attempt to focus on the four major topics of interest – internal variability, the NAO, nesting and ensemble approaches. We now open our introduction with the explanation of internal variability, introduce the NAO as a mode of internal climate variability, continue with the NAO quantification and representation in various climate models, introduce the ensemble approach (in order to assess NAO internal variability) and close with the necessity of regional climate models when analyzing NAO responses in heterogeneous regions.
Finally, we integrate the aforementioned topics to present our research question, which in turn leads to our four key questions. Among the key questions, we changed the order (switch (b) and (c)) as this seemed more consistent with the analyses in the study.
We also removed the information on interactions of the NAO with other modes of climate variability, as it is not crucial for the study. Instead, we included a short paragraph on the interactions between NAO and the East Atlantic/Scandinavian Pattern in the discussion section (see below, response to RC1 l. 326)

l. 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.
The “positive state” refers to the positive NAO state introduced in (discussion paper) lines 26-27. To improve the readability, we rephrased:

l. 25-28: “Its two states, positive and negative, are evoked by planetary wave-breaking in the polar front, leading to antagonistic pressure behavior of two centers over the North Atlantic: one located within the subtropical high pressure belt (“Azores High”, AH), the second in subpolar regions (“Icelandic Low”, IL) (Benedict et al., 2004).”

l. 31-33: “Compared to neutral conditions, the positive NAO state leads to warmer and moister winters in northern Europe, but cooler and drier conditions in the south, and vice versa in the negative state (e.g., Hurrell and Deser, 2009; Pokorná and Huth, 2015; Woollings et al., 2015).”

ll. 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.

Thank you! We removed the "table of contents".

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

We re-structured this paragraph. It now starts with a more detailed explanation of the GCM LE before moving to the RCM LE. The information on the domains covered by the regional model follows immediately after:

l. 91-93: "As described in Leduc et al. (2019), these 50 GCM members were dynamically downscaled using the Canadian Regional Climate Model version 5 (CRCM5 Large Ensemble, 0.11° spatial resolution) over two domains covering Europe and north-eastern North America."

l. 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 uncertainty. We included a short comparison between both model ensembles after the first mention of the CORDEX ensemble:

l. 96-103: "Comparing the internal variability of the CRCM5 members with the IMS of a subset of the multi-model EURO-CORDEX (Coordinated Regional climate Downscaling Experiment) ensemble regarding winter temperature and precipitation, von Trentini et al. (2019) showed that both ensemble spreads are of comparable magnitude. The CORDEX ensemble consists of several GCM-RCM combinations set up in a coordinated modelling framework and aims at evaluating uncertainty due to model configuration (Giorgi et al., 2009). The similarity of the single-model and multi-model spreads suggests that a large fraction of the CORDEX ensemble spread can be explained by internal variability, despite the fact that it was not explicitly sampled within the CORDEX framework (where most models provided a single simulation, von Trentini et al., 2019)."

ll. 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 a judgement of importance of modes. We rephrased the statement:

l. 103-104: "Therefore, the GCM and RCM ClimEx ensemble can be expected to capture the range of winter temperature and precipitation internal variability despite the set up with a single model."

ll. 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 changed 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 kept 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 also placed an explanation in the manuscript:

l. 114-115: "ERA-I variables t2m, tp and msl were chosen as they were assumed to most accurately represent the variables from the GCM and RCM models."

l. 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 removed it.

Following a suggestion of Referee Comment 2, we replaced Table 2 with a labeled map (Fig. 1) indicating the size and position of the regions of interest:

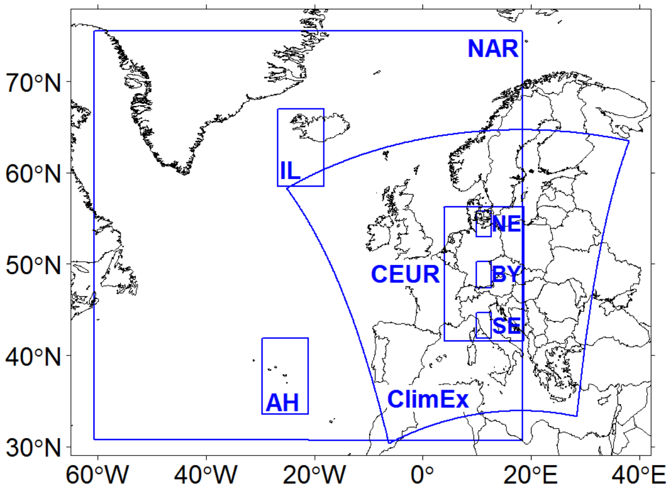

Figure 1: Regions of interest. Abbreviations and domain sizes in terms of GCM grid cells (2.8°) are as follows: AH – Azores High (3×3); IL – Icelandic Low (3×3); NAR – large-scale North Atlantic region (28×16); CEUR – Central Europe (5×5); NE – northern Europe (1); BY – Bavaria (1); SE – southern Europe (1); ClimEx – domain used in ClimEx project (extent approximately 22×12 after resampling to GCM grid).

l. 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. l. 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 now use past tense.

ll. 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, 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 rephrased the paragraph accordingly:

l. 156-159: "This study focused on inter-annual analyses which were conducted for two time horizons covering 30 years each. The chosen period length was assumed to include major fluctuations, like internal climate variations or several solar cycles, which might affect NAO phases (Andrews et al., 2015). Thus their influence can be assumed to be represented by the sampled NAO time series."

ll. 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.
We included a paragraph in the discussion section, which refers to this limitation.

l. 498-502: "It has to be added that this study evaluated two 30-year blocks rather than continuous time series, treating the NAO–response relationship as stationary during these blocks such that the IMS of both periods represents generalized conditions for 1981–2010 and 2070–2099. According to Comas-Bru and McDermott (2014), potential non-stationarity in NAO–response relationships can at least partly be attributed to influences of the EA/SCA patterns on the NAO, and especially the geographical position of the North Atlantic SLP gradient."

ll. 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 changed the wording:

l. 169-174: “Since the NAO is known to be strongest in winter (Hurrell and Deser, 2009) and the connection between station-based indices and NAO responses tends to be best in winter (see Pokorná and Huth, 2015, for months DJF), analyses were performed for this season only. Preliminary tests had shown that correlations and links between the NAO index and the climate variables were more distinct from noise, if March was included as well. That is why an extended winter season was used here (DJFM, see also Iles and Hegerl, 2017; Hurrell, 1995; Osborn, 2004).”

ll. 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 rephrased the paragraph:

l. 176-181: “The NAO index was derived from ERA-I and CanESM2-LE data, resulting in 1 REF and 50 GCM realizations. The NAO is quantified in this study with an index which is closest to a station based or zonally averaged index. This allowed obtaining an index in a large data set (50 members during hist and fut time horizons) at justifiable computational time. Other than indices based on PCA, this index does not represent a “pure” NAO pattern, i.e. the variability of North Atlantic SLP without any other teleconnection patterns like the East Atlantic Pattern (EA) and the Scandinavian Pattern (SCA) (Moore et al., 2013). Instead, it directly represents the winter SLP gradient over the North Atlantic.”

ll. 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 was not intended to present results. We included some more information to enhance its relevance.

l. 228-237: “Internal variability was understood as being represented by the oscillations around the long-term mean of the time series of a given variable (Hawkins and Sutton, 2011). In this point of view, IMS of the LE originates from the superposition of all 50 realizations with their respective inter-annual variability. As the climatic evolution of all 50 members is equally likely by construction of the ensemble, this spread represents an envelope of possible sequences of weather events at any given time step or location. This allows to sample internal variability at single points in time as the range of the members’ values.
Therefore, the NAO–response relationship was analyzed individually for each GCM and RCM member (as is done e.g. in Woollings et al., 2015).
Aggregations to ensemble means (like in Deser et al., 2017) and standard deviations (sd, see also Leduc et al., 2019; Déqué et al., 2007), the latter representing the IMS in maps, were only performed for illustrating purposes in order to avoid masking model internal variability (Zwiers and von Storch, 2004).”

l.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. We omitted the part “REF”. Two sentences later, the NAO index derived from ERA-I is defined as being “a reference” for the rest of the study.

l. 244: “First, a NAO index was calculated from the ERA-I reanalysis.
l. 246-247: “For further analyses it will therefore serve as a reference.”

ll. 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?
The sentence in the discussion paper 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 LE 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 different sample sizes (n_ERA-I = 30, n_CanESM2-LE =1500); the ERA-I sample is only one realization which is compared with the mean of 50 ensemble realizations, so deviations between the distributions may occur.
We changed (old) Fig. 1 (see Fig. 2), using CDFs rather than histograms as this removes the problem of binning and also represents the different sample sizes (see smooth curves for CanESM2 as opposed to steps for ERA-I)

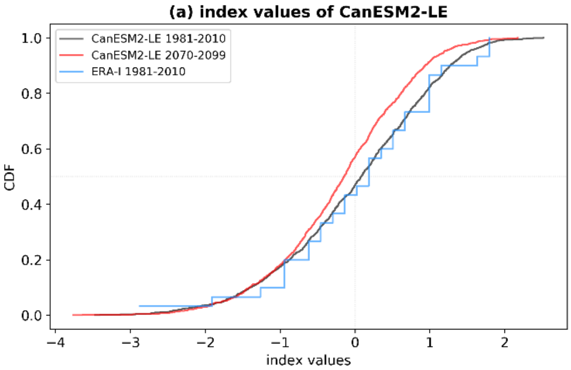

Figure 2: Cumulative density functions (CDFs) of NAO index values. (a) distribution of all CanESM2-LE (n = 50 × 30 per period) and ERA-I (n = 30) NAO index values. Black: 1981–2010 CanESM2-LE, red: 2070–2099 CanESM2-LE, blue: 1981–2010 ERA-I.

In addition, we complemented the corresponding paragraph:

l. 248-253: “The CanESM2-LE produces NAO index values which follow a distribution similar to the ERA-I data (centered over zero, slight surplus of low positive NAO values, see Fig. 3 (a)). The CanESM2-LE distribution appears smoother due to a larger sample size (n = 1500 for CanESM2-LE and n = 30 for ERA-I). Maximum and minimum index values (x-axis in Fig. 3 (a)) of some of the 50 members exceed those of the REF realization; thus, the REF realization lies well within the ensemble IMS. The future NAO index shows a similar distribution of values, but with slightly less positive and more negative values (red curve in Fig. 3 (a)).”

l. 214 *“…original data in***to** *three subsets…”*

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

We adopted both points, changing “Indifferent” to “neutral”.

l. 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, i.e. neutral SLP conditions. We changed the wording in both text and header (see Fig. 3 below).

ll. 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 (old Figs. 2a and 2d, respectively, see Fig. below). SLP over Greenland rises to about 1025 hPa in CanESM2-LE and about 1015 hPa in ERA-I data (hence overestimation in CanESM2-LE with respect to ERA-I; see yellow circles in Fig. 3); over the North Sea, SLP reaches 1000 hPa in the CanESM2-LE and 1010 hPa in ERA-I (hence underestimation in CanESM2-LE with respect to ERA-I; see red circles inserted in Fig. 3 below).

We clarified the wording and changed the coloring of the SLP maps to better visualize the differences.

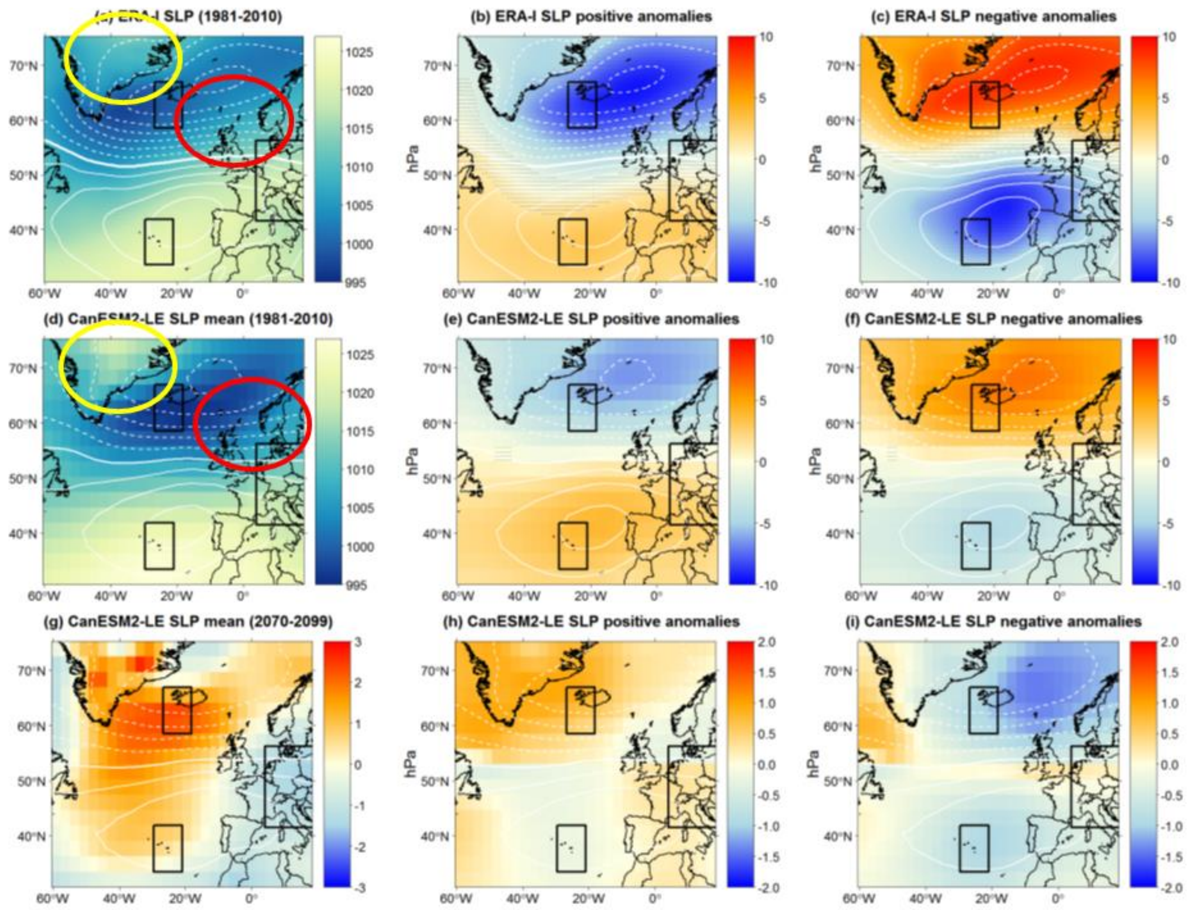

Figure 3: NAR winter mean SLP [hPa] composites in REF ((a)–(c)) and GCM ((d)–(i)) data showing long-term neutral conditions (left column), NAO positive (mid column) and negative anomalies (right column). (a)–(f): for 1981–2010. (g)–(i): 2070–2099 changes with respect to 1981–2010 in GCM data. White isolines: difference between positive and negative anomalies by a step of 2.50 hPa, as e.g. in Hurrell (1995), solid: positive, dashed: negative, bold line: zero. Stippling in subpanels (a)–(f): regions where the anomaly is smaller than the standard error of the composite samples. Black boxes: AH, IL and CEUR regions (see Fig. 1).

The correspondent paragraph says now:

l. 272-276: “Under neutral NAO conditions, the North Atlantic region is characterized by a pressure dipole. This structure is intensified and tilted clockwise in the CanESM2-LE ensemble mean (middle row of Fig. 4) compared to REF (top row). The mean SLP difference between the CanESM2-LE mean and REF reaches up to 10 hPa in both directions. SLP values are higher over Greenland and lower over the North Sea in the CanESM2-LE compared to ERA-I (compare subpanels (a), (d) in Fig. 4).”

l. 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 removed the “also”.

ll. 239-240 *“…the spatial patterns of ERA-I and CRCM5/ERA-I differ* **more** *strongly than in Fig. 3,...”*
We corrected the sentence.

l. 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. We rephrased this sentence:

l. 321-322: “In comparison to the neutral state, positive phases are also accompanied by more humid conditions in the north, and drier conditions in the south of the CEUR domain (see Fig. 8).”

l. 256 The NAO explains less variance than what?
l. 257 tas std decreases less than what?
l. 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.
ll. 259-260 Could you please give a figure reference for the claims made here?

In all three cases (l.s 256, 257, 259), we compared the historical and future time horizons. When rewriting (and shortening) the paragraph, we included the necessary references and the figure reference. However, the sentences regarding comments l.256 & 257 were removed from the manuscript as the information is already included in the previous sentences.

l. 335-338: "Future NAO–climate relationships weaken in general compared to the historical ones for all variables. The spatial patterns of NAO-induced change do not change considerably between both periods. The response to the NAO, α1, is clearly reduced in nSAT mean as is nSAT std, and there is also a reduction in PR sum change (panels (g), (i) in Figs. 6–8)".

l. 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 sd maps/subpanels (d), (f) in Figs. 6—8) and spatially aggregated in NE, BY, SE (see Fig. 10). We inserted a short note on this matter:

l.341-342: "The representation of internal variability in the GCM and RCM regarding the responses to the NAO in CEUR and subset regions NE, BY, SE is assessed via differences in the IMS of the CRCM5-LE compared to the CanESM2-LE."

l. 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 included scatterplots regarding the relationship between LE ensemble means and sd, showing also several correlation measures (rather than just correlations as suggested in the final response; see Fig. 4).
Since in this case it is of no importance whether α1 is positive or negative –we are interested in whether a strong response is related with a large inter-member spread – only absolute α1 ensemble mean values were used.
The scatterplots indicate that particularly for nSAT std, the relationship may be seen as linear (the two clusters in the GCM data are probably due to the small domain size). nSAT mean and PR sum do not show a clear linear relationship in the GCM data.
RCM data shows more linearly oriented point clouds during the historical period for both nSAT variables, but a decrease in the correlation during the future period, as well as lower ensemble mean and sd values for both nSAT variables. PR sum though shows some higher ensemble mean and sd values.

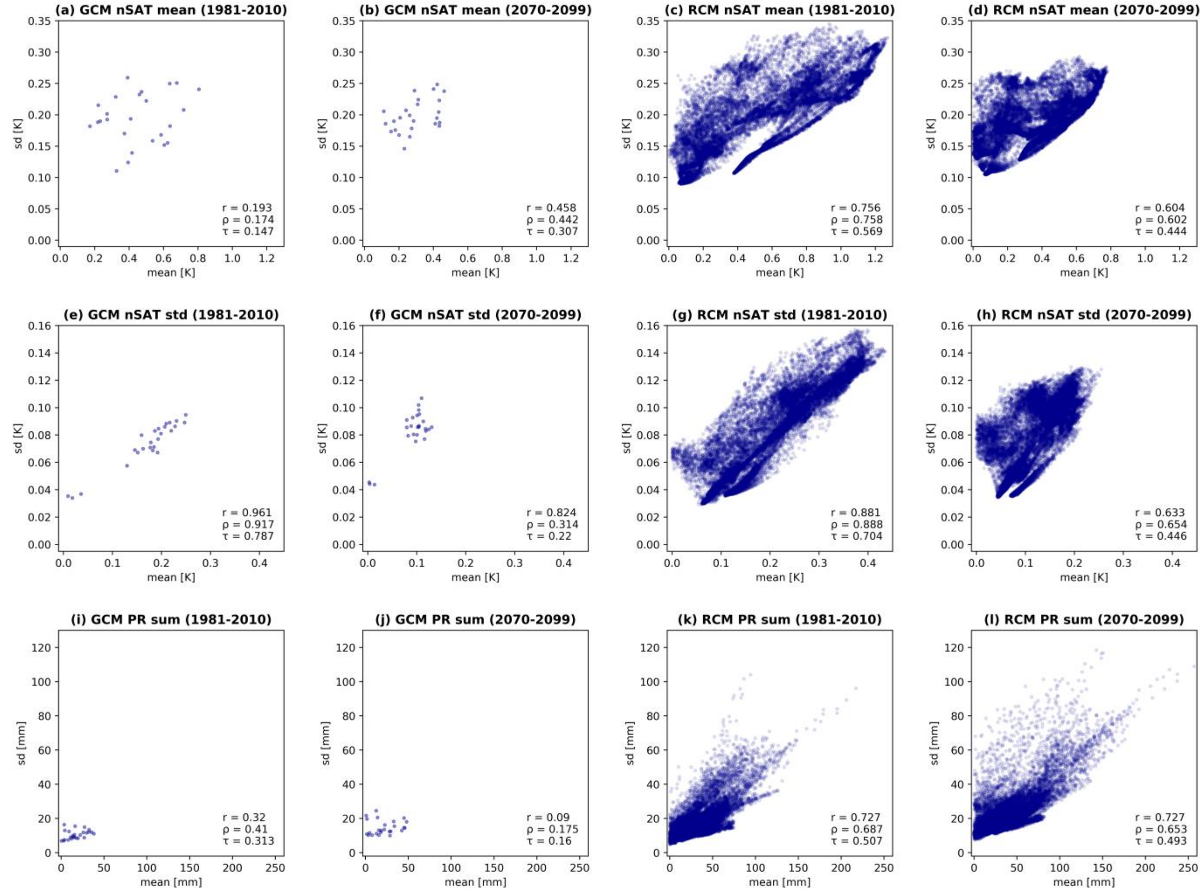

Figure 4 : Relationship between LE mean and sd values for variables nSAT mean (a)–(d), nSAT std (e)–(h), PR sum (i)–(l) for hist and fut time frames. Upper right corner: r – Pearson correlation coefficient, – Spearman rank correlation coefficient, – Kendall's Tau.

We also updated the corresponding paragraph:

l. 354-361: "Largest deviations for nSAT mean are found in continental regions of CEUR, but they do not simply correspond to high or low α1 (see also Fig. A3 (a)–(d)). Low IMS corresponds mostly to Alpine and sea regions. For nSAT mean, the signal-to-noise ratios (SNR) between ensemble mean and sd exceed 1 in most regions north of the Alps (see regions without stippling in Fig. 6). nSAT std shows SNR < 1 in the northern parts of the CanESM2-LE data (see Fig. 7 (c)) and in the Alpine region of the CRCM5-LE data (Fig. 7 (e)). This variable shows a strong linear relationship between LE mean and sd (Fig. A3 (e)–(h)). Regarding PR sum, RCM members vary most in regions with highest absolute α1 values and altitudes, but there is no clear dependence in GCM (Fig. A3 (i)–(l)). For PR sum, there is an east-west corridor of SNR values below 1 which accompanies rather low α1 values (see Fig. 8)."

l. 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). We changed the reference to "panels (h), (j) in Figs. 3-5". It now reads:

l. 361-363: "In addition to future changes in the NAO responses ensemble means, there is also a change in the spatial distribution of the IMS expressed as ensemble sd (see subpanels (h), (j) in Figs. 6—8)."

ll. 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.

We reordered the sentence:

l. 385-387: "Apart from PR sum in the NE region (both time horizons), no significant difference between the spread amplitudes of GCM and RCM is visible ($p \leq 0.05$, F-test)."

l. 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 nSAT mean, PR sum and nSAT std time series of the CRCM5 with the CanESM2 in the subset regions NE, BY, SE. Therefore, we correlated the time series of, e.g. nSAT 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 added an explanation to the manuscript.

l. 389-390: "To evaluate the co-variability of CanESM2 and CRCM5 data in the subset regions, time series of the response variables originating from both data sources were correlated member-wise (see Fyfe et al., 2017, for a similar approach)."

ll. 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 included a more detailed and revised paragraph regarding this issue.

l. 390-400: "As can be seen in Fig. 11, highest accordance on average is reached for nSAT mean in both periods, indicating that CanESM2-LE and CRCM5-LE show very similar temporal variability for this variable. The co-variability of GCM and RCM time series is weaker for PR sum and nSAT std than for nSAT mean in both periods. Also, the IMS is larger for PR sum and nSAT std than for nSAT mean. This finding suggests that there is a larger discrepancy in portraying PR sum and nSAT std in the RCM with respect to the GCM compared to nSAT mean. The correlations between CanESM2 and CRCM5 subset regions are in general significantly lower under future climate conditions compared to the historical ones, apart from nSAT mean in BY, PR sum in SE and nSAT std in BY (see text in Fig. 11). For nSAT std a shift of the distribution of r towards slightly larger values is visible. All variables exhibit a future IMS increase, though not all subset regions are affected (see e.g. nSAT mean BY or nSAT std SE in Fig. 11). This suggests that under future climate conditions a considerable reduction of GCM–RCM co-variability needs to be taken into account, at least for PR sum and (weaker) for nSAT mean. "

ll. 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 (old) Fig. 8. suggest that there is a larger discrepancy in portraying PR sum and nSAT std in the RCM with respect to the GCM than for nSAT mean. We addressed this issue when rewriting the entire paragraph (see response to previous comment).

ll. 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 removed the sentence.

l. 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 (old) Fig. 2 (d). The sentence was updated with the corresponding information:

l. 415-416: "The strong SLP gradient under neutral NAO conditions over the North Atlantic noted in the CanESM2-LE though suggests an overestimation of the local atmospheric circulation with too strong westerlies."

ll. 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. However, when rewriting the discussion, this sentence was removed as the information is already provided in the results section.

l. 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 conducted an uncertainty assessment of the samples for positive and negative NAO composites (referring to Fig. 2, panels (b)-(c) and (e)-(f)). Therefore, we estimated the standard error of the arithmetic mean on each grid cell for the ERA-I data and compared it with the CanESM2 samples (which are considerably larger). We included stippling in Fig. 2 (a-f, only visible in b-c, e-f) to show where the signal is larger than the standard error.

The results section was updated accordingly:

l. 276-285: "Long-term neutral states of both data sources show robust signals in the entire NAR region (i.e., no stippling). This suggests that the different patterns in GCM and REF data are not singularly artefacts arising from different sample sizes. The GCM multi-member composites of positive and negative phases show less pronounced SLP anomalies than the REF data. This difference between GCM and REF may be due to the fact that REF composites were derived from n = 3 negative and n = 4 positive years whereas the GCM data provided n = 264 negative and n = 263 positive years during 1981–2010. Regions with strong sampling uncertainties, i.e. where the standard error is larger than the anomaly, are indicated with stippling in panels (a)–(f). These regions are mostly found in the transition region between the wider ERA-I AH and IL nodes, whereas the SLP anomalies at the NAO centers of action show less uncertainty. The GCM patterns are more robustly assessed (i.e. less prone to sampling uncertainty) as can be seen by the very small area with stippling in which the sign of the anomaly may not be assessed robustly."

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 since there is only a single realization of observations.

l. 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.

Following this suggestion, we included 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).

l. 417-424: "Since the NAO index was obtained from raw SLP data, it contains the contribution of the NAO, but possibly also of micro-climatic noise or other teleconnection patterns like the East Atlantic (EA) and the Scandinavian Pattern (SCA) which interact with the NAO and exert a notable control on the North Atlantic SLP gradient (Moore et al., 2013). Moore et al. (2013) investigated the contributions of the North Atlantic teleconnections NAO, EA and SCA in reanalysis data by separating them with empirical orthogonal functions. The authors found that the NAO accounts for about one third of winter SLP variability, and the second and third leading modes for roughly 20 % and 15 %, respectively (see also Comas-Bru and McDermott, 2014). Thus the results presented here may be seen as representing the superposition of these atmospheric modes."

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

We rephrased the sentence:

l. 410-411: "The first is found in the here presented results, the latter is not clearly visible in the chosen domain."

ll. 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 removed it in the revised manuscript.

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

Thanks, we removed it.

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

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. We included a more detailed paragraph:

l. 456-469: "Apart from the coarser pattern resolution, there is also a shift in the spatial climate patterns in the CanESM2-LE within the CEUR domain with respect to ERA-I data which is not found in the CRCM5-LE: for example, typical continental climate features, such as high nSAT variability (as indicated by Fig. 6), are shifted southwards in the CanESM2-LE with respect to CRCM5-LE data (or ERA-I). This shift may be explained by the fact that due to coarser spatial resolution the GCM topography shows land grid cells where the Mediterranean or the Baltic Sea extend in ERA-I and CRCM5; thus, in the GCM, the continent Europe also occupies a region which is sea in ERA-I. Assuming that the land–sea distribution affects the climate evolution, the GCM also experiences a geographical shift of climatic characteristics (such as continental properties) compared with the ERA-I and RCM data within the study domain. Another example is the dividing line for NAO–PR sum relations (see Fig. 8) which shows a displacement in the GCM compared to the RCM. This displacement is related to the GCM orography which deviates due to the coarser spatial resolution in shape, position and height from the RCM orography. These findings suggest that similar responses of GCM and RCM to the NAO may not be visible at the same geographical location (i.e. coordinates), but under similar geographical conditions (exposition, altitude, distance to sea). Continuing this thought, the RCM reproducing the spatial climatic patterns in the "correct" location is another expression of the RCM added value for regional or local scale analyses. However, for general statements on this issue, analyses on a larger domain would be necessary."

l. 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. We included the correspondent information in the data section. The overestimation of average nSAT and PR does not affect the findings regarding the correlation coefficients since these are based on the changes/variability, rather than on background temperature/precipitation. However, the large discrepancies among local α1 between the GCM and RCM maps show a clear resemblance with the background nSAT and PR fields.

l. 118-124: "Figure A1 shows that the CRCM5 tends to underestimate (overestimate) mean winter nSAT mean in the northern (southern) part of the domain, regardless of the driving data (see first column for ERA-I and third column for CanESM2), whereas winter PR sums are overestimated in nearly the entire domain with strongest values in the south-eastern part. The GCM overestimates (underestimates) nSAT mean north (south) of the Alps. PR sum is underestimated in the entire domain apart from the western side of the Alps in the GCM. However, as this study will focus on changes in nSAT and PR induced by the NAO (see Section 2.2.4), biases are of no large relevance in general, but may show some influence when it comes to regions with particularly high PR sum values."

l. 327-332: "In the CRCM5-LE, single spots in mountainous regions (e.g. in the Dinaric Alps) show extremely high PR sum α1 values (up to ±220 mm per unit index change) where long-term mean PR sums are also very high. This stresses the more detailed production of geographical features, but also the tendency to evolve local extreme values in the high-resolution RCM (see similar results for local daily extreme precipitation in Leduc et al., 2019) which may even be noted in the (spatially aggregated) bias towards the GCM (see Fig. A1 (f))."

l. 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 rephrased the statement accordingly.

l. 478-480: "However, as the ensemble mean (GCM and RCM) reproduces patterns very similar to the observed ones, the atmospheric dynamics behind can be regarded as correctly reproduced in all members."

l. 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 (old) 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. When rewriting the discussion section, we removed this sentence, but the information about comparable characteristics is mentioned in the results and conclusions section, e.g.:

l. 267-268: "The members also show no systematic correlation with the REF NAO index despite similar statistics (see also Fig. 9)."

l. 538: "The ensemble also shows comparable climate statistics with the REF time series and patterns."

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

Thanks for this suggestion. However, this sentence was removed as its information is already implicit in section 3.2.1:

l. 354-361: "Largest deviations for nSAT mean are found in continental regions of CEUR, but they do not simply correspond to high or low $\alpha 1$ (see also Fig. A3 (a)–(d)). Low IMS corresponds mostly to Alpine and sea regions. For nSAT mean, the signal-to-noise ratios (SNR) between ensemble mean and sd exceed 1 in most regions north of the Alps (see regions without stippling in Fig. 6). nSAT std shows SNR < 1 in the northern parts of the CanESM2-LE data (see Fig. 7 (c)) and in the Alpine region of the CRCM5-LE data (Fig. 7 (e)). This variable shows a strong linear relationship between LE mean and sd (Fig. A3 (e)–(h)). Regarding PR sum, RCM members vary most in regions with highest absolute $\alpha 1$ values and altitudes, but there is no clear dependence in GCM (Fig. A3 (i)–(l)). For PR sum, there is an east-west corridor of SNR values below 1 which accompanies rather low correlation values (see Fig. 8)."

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

"Less" is referring to a comparison between historical and future time periods. We rephrased:

l. 527-529: "As less nSAT and PR variance is explained by the NAO in the future climate projections than in the historical period, the influence of this climate mode on CEUR climate may be seen as potentially reduced."

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 put the answers to key questions (a)-(d) in bullet points.

l. 535-552: "

(a) Both large ensembles within the ClimEx project climate model chain are able to depict a robust NAO pattern under current forcing conditions. Each member represents a distinct climate evolution while sharing comparable statistics with all other 49 realizations and producing NAO and response patterns that are more robust than patterns of single realizations. The ensemble also shows comparable climate statistics with the REF time series and patterns. **The clearly visible connection of the NAO with nSAT mean and PR sum follows well-known patterns. The influence of the NAO on nSAT variability, as expressed by the analyses on nSAT std, is also remarkable**.

(b) The RCM is able to reproduce the large-scale SLP pattern and realistic response patterns in the analyzed domain. Clearly more topographic features are visible in the CRCM5-LE than in the CanESM2-LE which **suggests** added value of the RCM regarding the evaluation of small-scale NAO impacts. Deviations of nSAT and PR responses between members vary spatially within the domain and are found mostly in regions with strongest NAO responses.

(c) Internal variability of the NAO pattern is expressed very well within the 50 member single-model ensemble, and easily spans the observations regarding various indicators. The range of NAO responses is represented consistently between the driving GCM and the nested RCM. The spread is shifted towards stronger NAO–nSAT/PR relations in the RCM compared to the GCM in both time horizons.

(d) Concerning climate change, several changes go hand in hand: the winter index variability is reduced, the overall winter variability of nSAT and PR and also the fraction of NAO-explained nSAT is reduced, the relationship between NAO and response variables is weakened, the RMS* error regarding the large-scale SLP pattern between GCM and RCM slightly increases, and the co-variability of CanESM2 and CRCM5 subset regions for all weather variables is reduced.

ll. 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 considered this suggestion. Please have a look at the response to the previous comment (bullet point (a), sentences in bold).

l. 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 rephrased to “suggests” (also in bold in response to the comment regarding the conclusions, bullet point (b)).

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

We adopted the suggestion.

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 corresponded to lines of equal correlations between the NAO index and the nSAT mean/nSAT 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 changed 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 6—8 gain more clarity in doing so. Also, headers and captions were adjusted. See the following Fig. 5 as an example (the same changes were applied to the plots for nSAT std and PR sum).

We also included the information from former appendix Figs. A3-4: regions with signal-to-noise ratio < 1 are indicated by stippling.

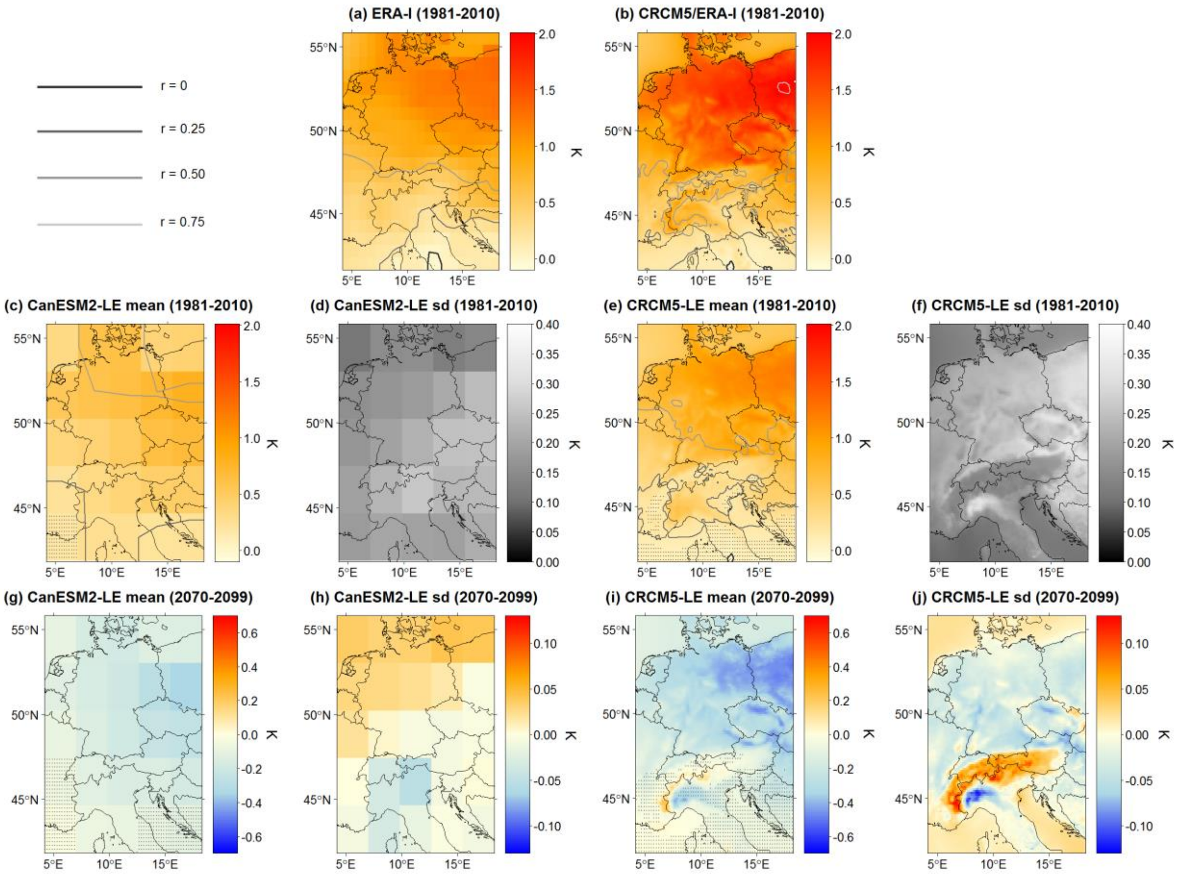

Figure 5: Spatial patterns of change in nSAT mean (in [K]) for a unit change in the NAO index for ERA-I, CRCM5/ERA-I, CanESM2-LE and CRCM5-LE in 1981–2010 ((a)–(f)) and the difference of 2070–2099 with respect to 1981–2010 ((g)–(j)). Both 50-member ensembles are represented with ensemble mean and sd representing the IMS. Grey lines in the ensemble mean maps represent the Pearson correlation between nSAT mean and the NAO index at an increment of 0:25; grey shadings see legend in upper left panel. Grey stippling in the ensemble mean maps show regions were SNR < 1, SNR being the signal-to-noise ratio between the 30 year ensemble mean and sd of GCM and RCM LEs in both time periods.

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 corrected the names:

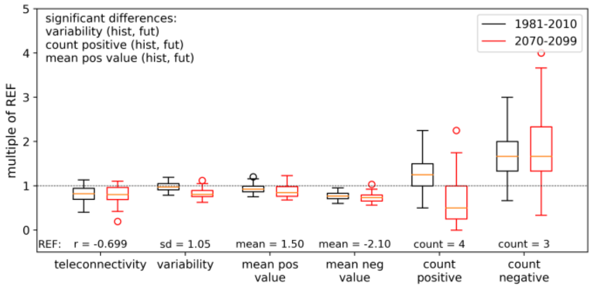

Figure 6: Several index statistics of all 50 CanESM2-LE members expressed as multiples of the respective ERA-I value (REF value set to 1.0): teleconnectivity (Pearson correlation between AH and IL time series), index variability (expressed as standard deviation in time of index time series), mean value of all positive (negative) phases and count of all positive (negative) phases in a single realization. Positive (negative) years are defined by an absolute index value exceeding 1. Text in upper left corner: significantly (p≤ 0.05, using an unpaired Mann-Whitney/U-test) different outcomes in the fut time frame.

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 nSAT mean (a), PR sum (b) and nSAT std (c). Thus “matching” refers to the same member in the GCM and RCM. We included a detailed description of the metric in the text (see also response to comment line l. 306), and changed the caption accordingly:

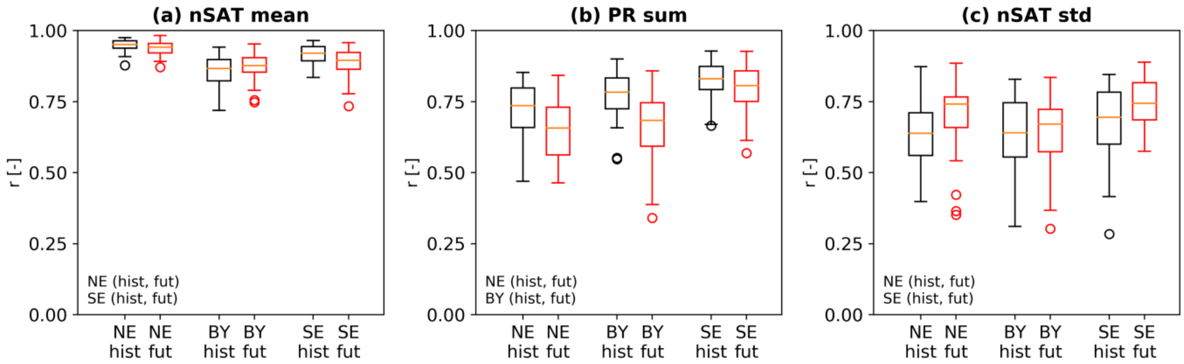

Figure 7: Temporal co-variability of CanESM2 and CRCM5 subset regions in all 50 members. Each boxplot represents 50 Pearson correlation coefficients between the time series of variables nSAT mean (a), PR sum (b) and nSAT std (c) in the subset regions of CanESM2 members and the corresponding CRCM5 members. Time periods used for correlations: 1981–2010 (hist, black), 2070–2099 (fut, red). For regions NE,BY, SE see Fig. 2. Text denotes combinations of which the differences are significant at $p \leq 0.05$ using an unpaired Mann-Whitney/U-test.

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)) andCRCMR5-LE ((e)-(f)) under historical and future conditions. We extended the caption accordingly:

“Ratio of nSAT α1 and winter mean daily standard deviation of nSAT for CRCM5/ERA-I (a) and ERA-I (b) under historical conditions and CanESM2-LE mean ((c)–(d)) and CRMC5-LE mean ((e)–(f)) under historical and future climate conditions. The panels show the fraction of nSAT 1 on winter mean daily standard deviation of nSAT.”

References in this response:

Andrews, M., Knight, J., Gray, L. (2015): A simulated lagged response of the North Atlantic Oscillation to the solar cycle over the period 1960–2009, Environmental Research Letters, 10.

Benedict, J., Lee, S., Feldstein, S. (2004): Synoptic View of the North Atlantic Oscillation, Journal of the Atmospheric Sciences, 61, 121–144.

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

Déqué, M., Rowell, D., Lüthi, D., Giorgi, F., Christensen, J., Rockel, B., Jacob, D., Kjellström, E., de Castro, M., van den Hurk, B. (2007): An intercomparison of regional climate simulations for Europe: assessing uncertainties in model projections, Climatic Change, 81, 53–70.

Deser, C., Hurrell, J., Phillips, A. (2017): The role of the North Atlantic Oscillation in European climate projections, Climate Dynamics, 49, 3141–3157.

Fyfe, J., Derksen, C., Mudryk, L., Flato, G., Santer, B., Swart, N., Molotch, N., Zhang, X.,Wan, H., Arora, V., Scinocca, J., Jiao, Y. (2017): Large near-term projected snowpack loss over the western United States, Nature Communications, 8.

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.

Hawkins, E., Sutton, R. (2011): The Potential to narrow uncertainty in projections of regional precipitation Change, Climate Dynamics, 37, 407–418.

Hurrell, J. W. (1995): Decadal Trends in the North Atlantic Oscillation: Regional Temperatures and Precipitation, Science, 269, 676–679.

Hurrell, J. W., Deser, C. (2009): North Atlantic climate variability: The role of the North Atlantic Oscillation, Journal of Marine Systems, 78, 28–41.

Iles, C., Hegerl, G. (2017): Role of the North Atlantic Oscillation in decadal temperature trends, Environmental Research Letters, 12.

Leduc, M., Mailhot, A., Frigon, A., Martel, J.-L., Ludwig, R., Brietzke, G., Giguère, M., Brissette, F., Turcotte, R., Braun, M., Scinocca, J. (2019): The ClimEx Project: A 50-Member Ensemble of Climate Change Projections at 12-km Resolution over Europe and Northeastern North America with the Canadian Regional Climate Model (CRCM5), Journal of Applied Meteorology and Climatology, 58, 663–693.

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

Osborn, T. (2004): Simulating the winter North Atlantic Oscillation: the roles of internal variability and greenhouse gas forcing, Climate Dynamics, 22, 605–623

Pokorná, L., Huth, R. (2015): Climate impacts of the NAO are sensitive to how the NAO is defined, Theoretical and Applied Climatology, 119, 639–652.

von Trentini, F., Leduc, M., Ludwig, R.(2019): Assessing natural variability in RCM signals: comparison of a multi model EURO-CORDEX ensemble with a 50-member single model large ensemble, Climate Dynamics, 53, 1963–1979.

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.

Zwiers, F., von Storch, H. (2004): On the role of statistics in climate research, International Journal of Climatology, 24, 665–680.

Responses to Referee Comment No. 2

In this study a regional climate model (CRCM5) is employed to dynamically downscale a single global climate model (CanESM2) large ensemble of climate change simulations to investigate the nature of downscaled responses to the modeled North Atlantic Oscillation (NAO) and its influence on future European climate. By employing a large ensemble, the authors are able to evaluate future downscaled responses associated NAO inter-annual variability in addition to mean changes. The authors set out four key questions related to, documenting the properties and fidelity of the modeled NAO in both the GCM and RCM; the associated screen temperature and precipitation responses in both models; and how such properties change under future external forcings (following the future CMIP5 pathway RCP8.5).

This is an interesting paper and ultimately worthy of publication. The authors present the problem from the perspective of downscaling teleconnections that exist in the driving data (ie the NAO). This is a subtle but critically important shift in focus for the dynamical downscaling community. The proper communication of teleconnection patterns/relationships from driving data to the RCM is essential for credible downscaled results. The use of a large GM/RCM ensemble pair positions the authors to say something definitive about this problem and offer guidance to the community.

The four key questions represent a clear and sensible plan for the paper. However, I found it difficult at times to cleanly connect a particular analysis performed by the authors with an answer to some of these questions. Specifically, I do not think that the authors addressed the first part of their question 3, "Do GCM NAO impulses propagate correctly into the RCM realizations" (l. 71). Perhaps a better way of stating this is, does the RCM faithfully represent the NAO pattern present in the driving data? This is a critical question in the authors' "model chain" (l. 65) that needs to be addressed before one moves on to evaluate the NAO responses. That is, if the largescale NAO pattern is not faithfully represented in the RCM domain in some location, the downscaled responses in that location would be less credible. The increased resolution and potentially improved physical processes present in the RCM themselves cannot correct the large-scale NAO pattern within the RCM domain. As the authors discuss, the NAO pattern is governed by "planetary wavebreaking in the polar front" (Benedict et al., 2004), which is intern influenced by external factors such as sea-ice, snow cover, sea-surface temperatures, ENSO, stratospheric circulation variability, solar variability, volcanic eruptions and the Quasi-Biennial Oscillation (eg Hall et al. 2014 https://doi.org/10.1002/joc.4121).

Given that the European domain is relatively small, and the experimental design employs spectral nudging in the RCM, the NAO pattern, and its interannual variability, should on balance be reasonably represented in the RCM. For the authors' stated plan, however, this needs to be verified. Given that the authors employ a large ensemble in their study, they are in the unique position to definitively address this issue and provide an example to the community of the type of analysis that is required to support the credibility of downscaled results in such complex problems. It is my recommendation that, prior to publication, the manuscript undergo major revision to address this issue and to improve its overall clarity. My detailed comments follow.

Thank you very much for this generally positive assessment of the study scope, but also for your concerns regarding key question 3. This question originally targeted the question whether the combination of NAO indices from the GCM and response variables from the RCM produces realistic looking NAO responses in the RCM. The suggested formulation changes its meaning towards the nesting of the NAO/SLP pattern itself. However, in light of the fact that indeed the assessment of large-scale SLP patterns in the RCM data is relevant but missing so far, this change of formulation is justifiable.

We adopted the suggestion in the major comment (see our point-by-point responses) and the ideas regarding different groups of correlations among member index time series.

Overall, we tried to optimize the structure of the paper, following the key questions raised at the end of the introduction in all following sections.

Major Comment:

RCM reproduction of NAO teleconnection in driving data

As part of the authors' model chain, it is essential to verify that the large-scale year-to year variations of the NAO pattern in surface pressure are faithfully reproduced (each year) in CRCM5 when driven by both ERA-I and CanESM2. Inspired by Fig. 2, the sort of analysis required would be as follows:

- interpolate monthly-mean timeseries of sea-level pressure (SLP) in the driving dataset onto the RCM grid (such interpolation is already done for the driving-data winds used for spectral nudging). Call this field SLP_Drive.

- take the difference of the RCM and driving data monthly-mean SLP on the RCM grid SLP_RCM - SLP_Drive, and then smooth the result retaining large scales that are representative of the driving data resolution:

D_m(i,j,t,n) = [SLP_RCM - SLP_Drive]_LRG

Here, i,j are lateral spatial coordinates of the RCM grid, t is time in units of years, n is ensemble member, and the subscript m corresponds to month (1-12). The smoothing operation, represented by the operator [ ]_LRG, can be performed with the same double-cosine transform used for the spectral nudging.

- derive a normalized root-mean-square difference map for extended winter, over the two 30-year periods displayed in Fig 2, over all ensemble members:

RMS (i,j) = Ave_(m=12,1-3){Ave_n { SQRT[Ave_t {D_m(i,j,t,n)ˆ2 }] / Var_Drive_m(i,j,n)}}

where, Ave_x is a simple averaging operators for the quantity x and Var_Drive_m is the variance in time of the driving data for each month and each ensemble member:

Var_Drive_m(i,j,n) = Ave_t {[SLP_Drive_m (i,j,t,n) - Ave_t {SLP_Drive_m (i,j,t,n) }]ˆ2}.

Normalization by Var_Drive_m is important as it indicates the size of an rms difference relative to the interannual variability in the NAO pattern at that location. Such an RMS map would provide a sensible measure of the difference in the driving data and RCM SLP patterns associated with the NAO, which need to be faithfully reproduced in each year. If RMS « 1 at a given location, then the large-scale NAO pattern is well represented there and one can conclude that the downscaling is consistently being performed on the "correct" large-scale flow. The larger RMS is, towards O(1) values, the more suspect the downscaled responses are at that location (ie a large-scale flow disconnected from the NAO in the driving data was being downscaled in these regions).
One should also do a significance test and indicate this by, say, filling in contours by color in only those regions that are significant at the 5% level. Given the size of the GCM/RCM ensemble, this should be quite robust (ie much of the canvas should be colored) and definitive statements could be made. This test would seem to be most well posed for the case of observational driving of the RCM (ie ERA-I driving of CRCM5 over the historical period 1981-2010). The large scales in that data are well observed and, because they came from the real system, they were influenced realistically by all processes and scales. Significant deviations in RMS(i,j) for ERA-I (ie RMS_ERA-I) would necessarily indicate a degradation of the NAO teleconnection in those regions of the CRCM5 domain.
If regions of NAO deviation in RMS_ERA-I were consistent with regions of NAO deviation in RMS_CanESM2 (in the historical and even the future periods), then this would indicate a systematic issue with the reproduction of the NAO pattern in the European domain in these locations and care should be taken in the interpretation of the downscaled responses in this, and possibly other RCM studies using the same domain.
Thanks for this very detailed suggestion! It is true that the original analysis did not include an assessment of the large-scale RCM SLP pattern. We adopted this suggestion with some slight modifications, the first one being that we interpolated the RCM data (and also the ERA-I driving data) to the GCM grid. This was done in order to not create additional errors during the interpolation onto the high resolution RCM grid. By aggregating the data, we also filtered the small scales, retaining only the large-scale patterns. We also decided to take the square-root of VarDrive. In this way, the RMS is dimensionless and may be interpreted as a root-mean squared difference. The RMS error was calculated on the entire ClimEx domain. Please see the next paragraph, which describes some insights gained from this measure:

l. 297-308: “Figure 5 maps the RMS* of the difference between driving data and RCM SLP during 1981–2010 for driving data ERA-I (a) and CanESM2 (b) and 2070–2099 (c). A value of RMS* ≥ 1 indicates that the root-mean squared error between the RCM and driving data is larger than the temporal variability in the driving data. In this case, the large-scale SLP pattern may not be seen as being correctly represented in the RCM data. The large-scale SLP pattern over the entire ClimEx domain, which also includes the CEUR, NE, BY and SE domains, is reasonably well represented: with RMS* < 1 in most parts of the entire ClimEx domain for both driving data sets and both periods (significant at p ≤ 0.05 using a t-test with a false detection rate < 0.1 to account for multiple hypothesis testing, see Wilks, 2016). All data sets show an RMS* increase towards the south, indicating that in these regions the control exerted by the lateral boundary conditions on the CRCM5 internal solution appears to be weaker. The RMS* is larger in the CanESM2/CRCM5 combination than in the ERA-I/CRCM5 combination, and slightly increases in the future period in the southern parts (see Fig. 5 (c)). In the CEUR domain (indicated as red box in Fig. 5), however, errors are low in general and therefore the NAO pattern of the driving data may be assumed to be correctly incorporated there.”

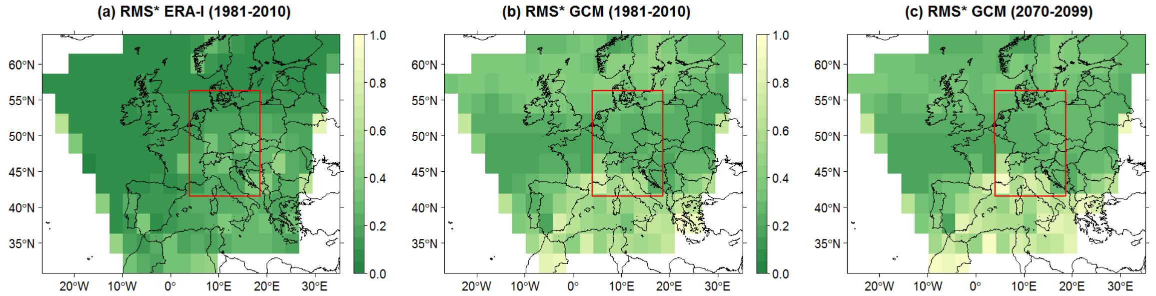

Figure 8: RMS of monthly SLP differences between driving data and CRCM5 members, calculated following Eq. (2). Colouring: RMS < 1 significant at p ≤ 0.05 with a false detection rate smaller than 0.1 (see Wilks, 2017). (a) for driving data set ERA-I (1981–2010), (b) for driving data set CanESM2 (1981–2010), (c) for driving data set CanESM2 (2070–2099). Red box: position of CEUR domain.

Minor Comments:

l.2 "natural variability". Later it seems, "internal variability" (l. 16) is used to refer to the same phenomenon. It would be helpful to be consistent throughout.

This is true. We note that terms like “natural variability”, “internal variability” and “noise” were used inconsistently in the study. We fixed this issue, using “internal variability” throughout the text.

l. 5-6. "its transfer from the driving model CanESM2 into the driven model CRCM5." Perhaps better wording might be "its representation in the driven model CRCM5 relative to the driving model CanESM2."

Thank you. However, when rewriting the abstract, this sentence was modified strongly. It now reads:

l. 5-7: “The overall goal of the study is to assess whether the range of NAO internal variability is represented consistently between the driving global climate model (GCM; the CanESM2) and the nested regional climate model (RCM; the CRCM5).”

l.11 "(b) impulses from the NAO in the CanESM2-LE produce" The use of the word impulses implies causality, which may be true for the one-way nesting/spectral nudging methodology but is not for the NAO itself. To avoid confusion perhaps say, "(b) reproduction of the CanESM2-LE NAO flow patterns in the CRCM5-LE produce"

We changed the wording in this sentence:

l.11-12: “Reproductions of the NAO flow patterns in the CanESM2-LE trigger responses in the high-resolution CRCM5-LE that are comparable with reference reanalysis data.”

l. 21 "is to apply slight differences in" -> "is to perturb"

ll.21-22 "with similar long-term climate statistics" This refers to a response rather than an experimental setup. I think it might be more correct to say "under identical external forcings"

Thanks, we adopted the wording suggestions:

l. 54:-56: “One way to trigger internal variability in GCM simulations is to perturb the initial conditions of the model, leading to several realizations of weather sequences under identical external forcing which also allow to derive a robust distribution of NAO index values.”

l. 44 "its dynamics in a future climate" -> "its fidelity in a future climate"

We changed the sentence accordingly:

l. 42-43: “While the typical NAO pattern and its impacts are usually correctly reproduced in global climate models (GCMs) (Stephenson et al., 2006; Ulbrich and Christoph, 1999; Reintges et al., 2017), its fidelity in a future climate remains uncertain.”

l.61 "is transferred correctly from the driving GCM into the driven RCM". Inter-member spread is not "transferred" from the driving model to the RCM. It would be clearer to say, "is represented consistently between the driving GCM and the driven RCM". Also, from my major comment, representation of NAO inter-member spread is a necessary condition from credible downscaled responses.

Thanks for your explanation. We rephrased the paragraph:

l. 67-69: “This study also targets the question, how global circulation variability, in this case the NAO teleconnection, affects local climate characteristics when downscaled using an RCM. It specifically aims at evaluating whether the range of internal variability is represented consistently between the driving GCM and the driven RCM.”

ll. 65-66 "finding robust NAO patterns which exceed the uncertainty due to internal variability in the ensemble." The phrase, "exceed the uncertainty due to internal variability” is confusing in this context. Perhaps say, "finding robust NAO patterns by significantly reducing sampling uncertainty associated with internal variability"

The suggested formulation is certainly clearer than the original one. We adopted the suggestion.

l.64-66: “The combination of the driving GCM and nested RCM large ensembles (LE) allows for analyzing the spread of NAO states and responses within one model chain, thus establishing the range of internal variability of the NAO, and finding robust NAO patterns by significantly reducing uncertainty associated with internal variability in the ensemble.”

l. 71 "Do GCM NAO impulses propagate correctly into the RCM realizations" perhaps better stated as, "Does the RCM correctly represent the NAO pattern present in the driving data" (ie my major comment)

We agree that (old) key question (c) is better stated in this way as the suggested wording also encompasses the additional analyses regarding the large-scale SLP pattern. We changed the order of the key questions, such that new key question (b) now reads:

l. 74-75: “Nesting approach: Does the RCM correctly incorporate the NAO pattern present in the driving data and produce realistic response patterns?”

ll. 68-74. These are excellent focal points/topics for the paper. It would be very helpful if these were better referred back to in the analysis, discussion, and summary sections so the reader can more easily keep track of which of these you are addressing and what progress you have made on each.

Thanks! We tried to structure the following sections accordingly. In the results section, the presentation of the results appeared to be easier when using a slightly different structure, i.e. the NAO index and spatial patterns (referring to key question (a) and partly (b)) is the first block, whereas the second block refers to the multi-model ensemble and internal variability (mostly key question (c)). (d) was found to be best presented alongside with the “hist” results.
In the discussion section however, results were integrated into the same structure as given by the key questions.

ll.101-103. two names are presented for each of three variables (eg msl/psl, t2m/tas, and tp/pr). I did not see a reason for this. If there is a reason it should be stated. If there isn’t, then it would be clearer if just one name was presented for each and used throughout the paper.

Thanks for this note. Please have also a look at the responses to Referee Comment 1 where we address a similar issue. The two names refer to different model variable output names (e.g. msl, t2m, tp were derived from ERA-I, psl, tas, pr from CanESM2 and CRCM5). We changed the analysis variable names in the following way psl → SLP, tas mean/std → nSAT mean/std, pr sum → PR sum. A short explanation was placed in the manuscript:

l. 114-115: “ERA-I variables t2m, tp and msl were chosen as they were assumed to most accurately represent the variables from the GCM and RCM models.”

ll.120-139. It would be very helpful here to provide a schematic, say of the range/extent displayed in Fig.2, where the RCM domain is indicated and where all of the regions discussed in this section were labeled . Not until I got to Fig 2 did the layout of things become clearer to me. Even then I had to look up Leduc (2019) to understand the relative positioning of the RCM domain.

We replaced Table 2 with a map showing all domains employed in the study. We think this is a more intuitive way to illustrate the position and extent of the domains than listing the boundary coordinates.

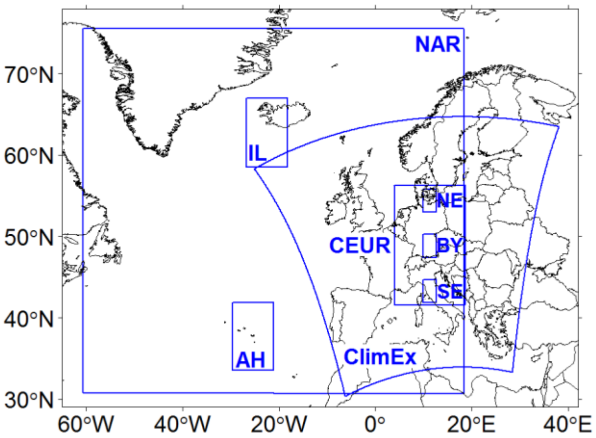

Figure 9: Regions of interest. Abbreviations and domain sizes in terms of GCM grid cells (2.8°) are as follows: AH – Azores High (3×3); IL – Icelandic Low (3×3); NAR – large-scale North Atlantic region (28×16); CEUR – Central Europe (5×5); NE – northern Europe (1); BY – Bavaria (1); SE – southern Europe (1); ClimEx – domain used in ClimEx project (extent approximately 22×12 after resampling to GCM grid).

l. 200 Fig.1 This figure is very faint and it is very hard to distinguish between the three cases being presented here. The authors should work on making these results clearer by using more vivid colours and/or fills.

Please have a look at the response to the next comment.

l. 208 "Pairwise correlations between the members". As discussed in Leduc et al. (2019), The CanESM2-LE was spawned in 1950 from 5 independent historical realizations (separated by 150 years of coupled integration each - including 50 years of preindustrial simulation between the launch of each ensemble member). As such, each of the 5 groups of 10 are highly independent of each other. The question of independence applies to the members within each group of 10 which has only 30years of coupled integration to develop independence prior to the 1981-2010 analysis period. Wouldn't a better check of independence be to form two correlation groups? The first would involve pairwise correlations between each member and the 40 other members from the 4 other groups that were spawned from a different CanESM2 realization in 1950. This first group would form a control assumed to be highly independent. The second group would involve pairwise correlations between each member and the 9 other members of the same group spawned from the same CanESM2 realization in 1950. Plots like figure 1b for this latter group could be compared to similar plots of the control group to assess the independence of the ensemble members most likely to have residual correlations during the 1981-2010 period.

This is a very nice idea. We performed an analysis following these steps. In order to better discriminate the different groups (and periods) we also switched from histograms to CDFs. Names of the two groups are SOIC – "same ocean initial conditions" (looking at members from the same family), and MOIC – "mixed ocean initial conditions" (looking at members from different ocean families) following Leduc et al. 2019. Also, we think that the colored lines are easier to read than histograms.

Referring to the independence of the members, we also included a new figure in the manuscript (see Fig. 11 below), showing the spreads among ten 5-member groups (see Leduc et al. 2019) for a daily NAO index in SOIC and MOIC groups. Groups from SOIC "start" with no standard deviations among the members. The spreads among members show no systematic differences after about a month after initialization.

A similar figure with winter NAO indices shows no differences in the entire period between the spreads among SOIC and MOIC members (Fig. 12 below).

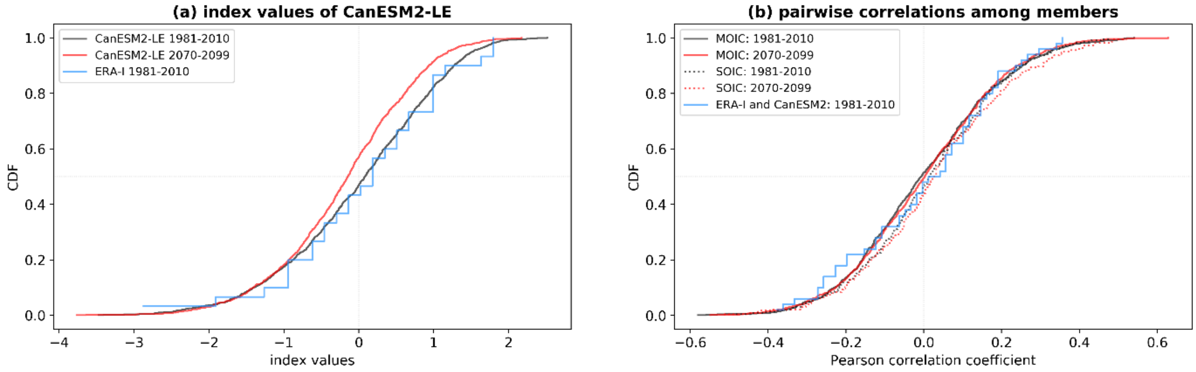

Figure 10: Cumulative density functions (CDFs) of NAO index values. (a) distribution of all CanESM2-LE (n = 50 × 30 per period) and ERA-I (n = 30) NAO index values. (b) pairwise correlations among member NAO index time series from the same ocean families (SOIC – same ocean initial conditions, dotted lines, n = 225), from different ocean families (MOIC – mixed ocean initial conditions, solid lines, n = 1000) and between ERA-I and all CanESM2 members (n = 50). Black: 1981–2010 CanESM2-LE, red: 2070–2099 CanESM2-LE, blue: 1981–2010 ERA-I.

The text says:

l.254-269: “For further analyses on the IMS as a measure of internal variability, the independence of the 50 ensemble members is of high importance. To investigate independence among the ensemble members in both 30 year time frames, it seems favourable to analyse pairwise member correlations. Although zero correlations do not automatically imply independence, clear correlations among embers would contradict the assumption of independence. In order to take into account the two perturbations during the production of the LE (1850 for 5 ocean families, 1950 for perturbations leading to 10 members per ocean family), these correlations were split in two groups like in Leduc et al. (2019): (i) correlations among the 10 members from the same ocean family (same ocean initial conditions in 1950, SOIC, n = 225, see dotted lines in Fig. 3(b)) and (ii) correlations between each member and the 40 members from the 4 other ocean families (mixed ocean initial conditions, MOIC, n = 1000, see solid lines in Fig. 3 (b)).
These correlations approximately follow a normal distribution with mu = 0. There is a slight surmount of low positive correlations in the SOIC group compared to the MOIC group which is (not significantly) stronger in the fut time horizon (see red and black dotted lines in Fig. 265 3 (b)). In general, the members are thus not seen as being dependent. As will be discussed below, the SLP pattern over the North Atlantic changes slightly in the future period. So the direct comparison between historical and future SOIC and MOIC correlations remains difficult. The members also show no systematic correlation with the REF NAO index despite similar statistics (see also Fig. 9). Thus, the ERA-I and GCM indices can be seen as not dependent realizations drawn from the same distribution.”

Regarding member independence, we placed a comment saying:

l. 89-91: “Regarding the atmospheric circulation, Fig. 1 shows that owing to the chaotic nature of the atmospheric system the daily NAO 90 index seems to lose dependence from the initial conditions within the course of one month after initialization (see Leduc et al., 2019, for a similar presentation of member independence).”

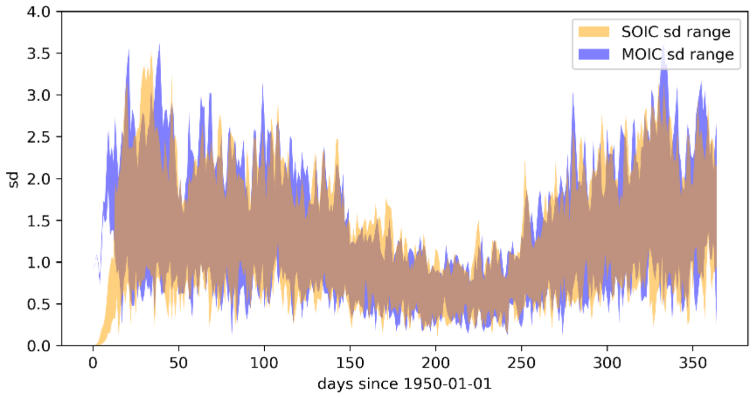

Figure 11: Inter-member standard deviation of a daily NAO index in the CanESM2-LE starting on 1 Jan 1950 as a function of time. The inter-member standard deviation was derived from ten groups of five members with the same ocean initial conditions (SOIC) and ten groups of five members with mixed ocean initial conditions (MOIC, following an approach in Leduc et al.,2019).

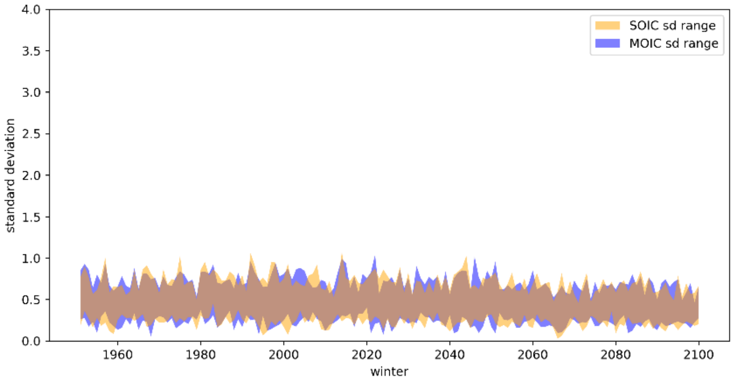

Figure 12: Like Fig. 11, but with winter (DJFM) NAO indices for the entire simulation period (1950-2100). This figure is not included in the paper.

l. 211 "They are not systematically related to the ERA-I (the nu201creferencenu201d) realization." Why would they be? I don’t understand the reasoning behind this correlation. If you are looking for a control group, a much larger group could be formed by the suggestion immediately preceding this point.

When correlating the ERA-I realization with the 50 CanESM2 members we were not so much looking for a control group. The idea was to evaluate whether ERA-I may show dependence with the CanESM2 members (i.e. non-zero correlations). We rephrased the sentence:

l. 267-268: “The members also show no systematic correlation with the REF NAO index despite similar statistics (see also Fig. 9).”

l. 214 "positive, negative and indifferent index values" -> "positive, negative and neutral index values"
l. 223 "it backs the choice" -> "it supports the choice"

We changed “indifferent NAO” values to “neutral” values throughout the manuscript, and also adopted the second suggestion.

ll.312-390 Discussion section. The references and discussion here are quite detailed and require constant back-and-forth reference to the earlier sections. For example, the opening statement of the second paragraph states, "The strong psl gradient suggests an overestimation of the local atmospheric circulation with too strong westerlies over the North Atlantic in the background state within the CanESM2-LE." What gradient? Where? The reader has to stop to review the previous sections to determine the context of this statement. This extends to the use of quantities that were defined in previous sections. For example, "Concerning NAO responses, they are most reliable in regions where r is significant (i.e. |r| > 0.361 for p nu2264 0.05,...". "r" may have been define earlier but the reader must stop here to find where that was to understand this context. (Also "Historical nu03b11 values" l. 327.) This discussion needs to be elevated somewhat out of the details of the previous section, summarize those outcomes and their implications, and connect back to the 4 key issues outlined in the introduction.

We agree that there is a lot "back and forth" which is related to the fact that we tried to respect the strict separation of the results and discussion sections. In order to improve readability though and following some comments of Referee Comment 1, we restructured the discussion section with respect to the four key questions (see section headers).

l. 321 "less prone to incidental fluctuations of single realizations" -> "less prone to sampling uncertainty"

Thanks, we adopted the correction.

ll. 323-325 "On the other hand, lower correlation values (|r| < 0.361) suggest that climate variability at the local scale evolves differently from the global teleconnection. In these cases, the NAO is not the most important contributor and nu03b5Y in Eq. (2) is dominant. Since the index was obtained from raw psl data, it contains the NAO contribution, but possibly also of other teleconnection patterns and noise." There is also the possibility that the large-scale NAO pattern in these regions was not reproduced correctly in the RCM. See my major comment.

ll 341-343 "Another possible explanation could be that the control exerted by CanESM2 through the CRCM5 lateral boundary conditions (LBC) is insufficient, but this is unlikely given the relatively small CRCM5 domain". Adopting the suggestion in my major comment would explicitly address this key issue.

We noted that in general the error is well below 1 in our CEUR domain, but especially southern regions exhibit a higher error compared to the rest of the entire ClimEx domain. This was also included in our discussion section:

l. 450-453: "Nevertheless, the influence of the lateral boundary conditions appears to vary over the CRCM5 domain, being a bit weaker in the southern part. It is worth noting that this feature is less pronounced when CRCM5 is driven by ERA-I as compared with CanESM2, highlighting the importance to investigate further the interactions between global atmospheric circulation, surface forcings (e.g. topography and land-sea contrasts) and local feedbacks."

<u>References in this response:</u>

Leduc, M., Mailhot, A., Frigon, A., Martel, J.-L., Ludwig, R., Brietzke, G., Giguère, M., Brissette, F., Turcotte, R., Braun, M., and Scinocca, J. (2019): The ClimEx Project: A 50-Member Ensemble of Climate Change Projections at 12-km Resolution over Europe and Northeastern North America with the Canadian Regional Climate Model (CRCM5), Journal of Applied Meteorology and Climatology, 58, 663–693.

Reintges, A., Latif, M., Park, W. (2017): Sub-decadal North Atlantic Oscillation variability in observations and the Kiel Climate Model, Climate Dynamics, 48, 3475–3487.

Stephenson, D., Pavan, V., Collins, M., Junge, M., Quadrelli, R. (2006): North Atlantic Oscillation response to transient greenhouse gas forcing and the impact on European winter climate: a CMIP2 multi-model assessment, Climate Dynamics, 27, 401–420.

Ulbrich, U., Christoph, M.(1999): A shift of the NAO and increasing storm track activity over Europe due to anthropogenic greenhouse gas forcing, Climate Dynamics, 15, 551–559.

Wilks, D. S. (2016): "The stippling shows statistically significant grid points": How Research Results are Routinely Overstated and Overinterpreted, and What to Do about I, Bulletin of the American Meteorological Society, 97, 2263-2273.

Responses to Editor Comment

…

I had the impression that discrepancies between the model and reanalysis data seem to be interpreted in two different ways. While you argue that the difference in figure 1 can be explained by the smaller sample size in the reanalysis (response to reviewer 1, ll. 205-206), in a different response to reviewer 1 (ll. 216-217) you state that the higher MSLP over Greenland in CanESM5 compared to observations points to an overestimation in the model. Would the larger number of realisations in the model be an alternative explanation for this discrepancy as well?

Thank you for your question. In both cases, the mean SLP and the anomalies, the size of the ensemble (or composite sample) may lead to differences. When looking at the anomaly composites though, this effect may be more relevant as the ERA-I anomaly composites have sample sizes n = 3 and n = 4 (negative and positive, respectively), whereas the long term mean ERA-I map has sample size n = 30 and may be thus somewhat more robust.

Please also consider the comments by reviewer 1 on lines 320 and 378 and your responses in this context and make sure to explain your interpretation and reasoning to the readers.

We tried to include the explanations to the referee comments where possible. Please have a look on the manuscript paragraph regarding line 320 below as an example:

l. 276-284: “Long-term neutral states of both data sources show robust signals in the entire NAR region (i.e., no stippling). This suggests that the different patterns in GCM and REF data are not singularly artefacts arising from different sample sizes. The GCM multi-member composites of positive and negative phases show less pronounced SLP anomalies than the REF data. This difference between GCM and REF may be due to the fact that REF composites were derived from n = 3 negative and n = 4 positive years whereas the GCM data provided n = 264 negative and n=263 positive years during 1981–2010. Regions with strong sampling uncertainties, i.e., where the standard error is larger than the anomaly, are indicated with stippling in panels (a)–(f). These regions are mostly found in the transition region between the wider ERA-I AH and IL nodes, whereas the SLP anomalies at the NAO centers of action show less uncertainty. The GCM patterns are more robustly assessed (i.e., less prone to sampling uncertainty) as can be seen by the very small area with stippling in which the sign of the anomaly may not be assessed robustly.”

Please make sure that definitions related to internal variability are explained and used consistently to avoid ambiguity. Currently, some phrases could be unclear (as pointed out by both reviewers). Some example are:
line 2 'natural variability'
lines 191/192: 'internal model noise', 'spread of internal variability'
…

This is true; the use of definitions related to internal variability was often quite ambiguous. In order to clarify, we now use “internal variability“ throughout the text and avoid the use of ambiguous wording like “internal model noise” etc. where possible. Additionally, we dropped the name “std.dev50” and stick to “inter-member spread” (IMS) when referring to the ensemble spread (maximum to minimum) and use “ensemble sd” specifically when referring to the maps of ensemble standard deviations.

Using a nested single-model large ensemble to assess the internal variability of the North Atlantic Oscillation and its climatic implications for Central Europe

Andrea Böhnisch[1], Ralf Ludwig[1], and Martin Leduc[2, 3]

[1]Department of Geography, Ludwig-Maximilians-Universität München, Munich, Germany
[2]Ouranos, Montréal,Québec, Canada
[3]Centre ESCER, Université du Québec à Montréal, Montréal, Québec, Canada

Correspondence: Andrea Böhnisch (a.boehnisch@lmu.de)

Abstract. Central European weather and climate is closely related to atmospheric mass advection triggered by the North Atlantic Oscillation (NAO), which is a relevant index for quantifying ~~natural~~ internal climate variability on multi-annual time scales. It remains unclear, though, how large-scale circulation variability affects local climate characteristics when downscaled using a regional climate model. In this study, 50 members of a single-model initial-condition large ensemble (LE) ~~(www.climex-project.org)~~ of a nested regional climate model are analyzed for a ~~climate–NAO relationship, especially its inter-member spread and its transfer from the driving model~~ NAO–climate relationship. The overall goal of the study is to assess whether the range of NAO internal variability is represented consistently between the driving global climate model (GCM; the CanESM2~~into the driven model~~) and the nested regional climate model (RCM; the CRCM5~~. The NAO pressure dipole is quantified in the CanESM2-LE by an extended station-based index; responses~~). Responses of mean surface air temperature and total precipitation to changes in the index value are ~~determined~~ expressed for a Central European domain ~~(CEUR)~~ in both the CanESM2-LE and CRCM5-LE ~~. NAO–response relationships are expressed~~ via Pearson correlation coefficients ~~(strength)~~ and the change per unit index change for historical (1981–2010) and future (2070–2099) winters. Results show that ~~(a)~~ statistically robust NAO patterns are found in the CanESM2-LE under current forcing conditions~~and (b) impulses from the NAO~~. Reproductions of the NAO flow patterns in the CanESM2-LE ~~produce correct~~ trigger responses in the high-resolution CRCM5-LE ~~. Relationships~~ that are comparable with reference reanalysis data. NAO–response relationships weaken in the future period, but ~~the amplitude of~~ their inter-member spread shows no significant change. ~~Among others, the results~~ The results stress the importance of single-model ensembles for the evaluation of internal variability. They also strengthen the validity of the ~~climate module in the ClimEx model chain~~ nested ensemble for further impact modelling ~~and stress the importance of single-model ensembles for evaluating internal variability~~using RCM data only, since important large-scale teleconnections present in the driving GCM propagate properly to the fine scale dynamics in the RCM.

1 Introduction

One of the major sources of uncertainty regarding short-term future climate projections is internal climate variability, while model climate response and greenhouse gases concentrations scenarios become more important sources of uncertainty on a longer-term time horizon (Hawkins and Sutton, 2009, 2011). The term internal variability denotes variability which is not forced by external processes (either anthropogenic or natural), but arises from the chaotic properties of the climate system itself (Leduc et al., 2019; Deser et al., 2012)~~. One way to trigger internal variability in global climate model (GCM) simulations is to apply slight differences in the initial conditions of the model, leading to~~ , i.e. from varying sequences of weather events ~~with similar long-term climate statistics. Global atmospheric modes of variability alter the~~ under identical external forcings. These sequences of weather events may be altered by global atmospheric modes of variability through the linking between large-scale circulation and local weather characteristics (like surface air temperature and precipitation). ~~They~~ Atmospheric modes can thereby establish periods of discernible states on multi-annual time scales.

Among these modes, the North Atlantic Oscillation (NAO) is particularly important for northern hemisphere climate. Its two states, positive and negative, are ~~accompanied by a stronger and weaker pressure gradient, respectively, over the North Atlantic region (Hurrell and Deser, 2009) . They are~~ evoked by planetary wave-breaking in the polar front, leading to antagonistic pressure behaviour of two centres over the North Atlantic: one located within the subtropical high pressure belt ("Azores High", AH), the second in subpolar regions ("Icelandic Low", IL) (Benedict et al., 2004). The resulting pressure gradient, which is stronger during positive and weaker during negative phases, affects large-scale extra-tropical circulation, especially the strength and position of mid-latitude westerly winds connected to the jet stream, and air mass advection during boreal winter (Deser et al., 2017; Hurrell and Deser, 2009). Compared to neutral conditions, the positive NAO state leads to warmer and moister winters in northern Europe, but cooler and drier conditions in the south, and vice versa in the negative state (e.g., Hurrell and Deser, 2009; Pokorná and Huth, 2015; Woollings et al., 2015) .

Commonly, the NAO is quantified with an index ~~making use of this pressure~~ that makes use of the air pressure or geopotential height gradient between AH and IL. The index may be calculated as a normalized difference of ~~sea level pressure or geopotential height~~ station measurements, spatially averaged ~~pressure~~ values of pre-set regions, or the region of highest ~~pressure~~ variance is obtained by principal component analysis (PCA) (Pokorná and Huth, 2015; Hurrell and Deser, 2009; Stephenson et al., 2006; Hurrell, 1995; Rogers, 1984). Each method has its advantages and limitations~~(see e.g. Pokorná and Huth, 2015, for a detailed survey of vari~~ ~~The positive state leads to warmer and moister winters in Northern Europe, but cooler and drier conditions in the South, and vice versa in the negative state (e.g., Hurrell and Deser, 2009; Pokorná and Huth, 2015; Woollings et al., 2015) . Several feedbacks with~~ . For example, station-based or fixed in space indices do not reproduce shifting NAO patterns and may be affected by micro-climatic noise and other teleconnection patterns ~~, like El Nio/Southern Oscillation, and with the ocean circulation are reported (Hurrell and Deser, 2009; Moore et al., 2013) . Further on , high pressure regions over Greenland (Greenland blocking) are likely related to the emergence of negative NAO phases (Hanna et al., 2015) .~~ (Hurrell and Deser, 2009; Osborn, 2004) . Indices based on PCA on the other hand are dependent on the chosen data domain for calculation and on the data set itself (Osborn, 2004) . The different approaches though lead to highly similar index time series (see e.g., Pokorná and Huth, 2015, for a detailed s

While the typical NAO pattern and its impacts are usually correctly reproduced in ~~GCMs~~ global climate models (GCMs) (Stephenson et al., 2006; Ulbrich and Christoph, 1999; Reintges et al., 2017), its ~~dynamics~~ fidelity in a future climate ~~remain~~ remains uncertain: the NAO is found as intensifying, but also counteracting global warming in the northern hemisphere ~~(“global warming hiatus”) (Iles and Hegerl, 2017; Deser et al., 2017; Delworth et al., 2016)~~ (“global warming hiatus”, Iles and Hegerl, 2017; Deser Similarly, the findings regarding the prevalence of future positive or negative states lack unity: ~~Analyses~~ Some analyses of CMIP5 models, for example, suggest ~~an increase of~~ more positive phases under rising greenhouse gas concentrations until 2100 ~~(e.g., Woollings et al., 2010; Kirtman et al., 2013; Christensen et al., 2013) . On the other hand, due to reduced sea ice extents which seem to be occasionally coupled with Greenland blocking, negative NAO phases may become more likely in future climate (Gillett and Fyfe, 2013; Hanna et al., 2015)~~ (e.g., Kirtman et al., 2013; Christensen et al., 2013) , others favour an increase of negative phases (Cattiaux et al., 2013) .

~~It is common in most of the mentioned studies~~ In most of these studies it was common to rely on one simulation per model and estimate ~~its performance~~ the model’s performance regarding the NAO by this single run. This approach allows for comparing different models (and observations). However, it is not possible to ~~directly~~ robustly evaluate the range of ~~the model internal NAO variability~~NAO index values, or whether the chosen simulation is a good representation of how this model simulates the phenomenon in question (Leduc et al., 2019). ~~Single realizations~~ Relying on single realizations possibly deteriorates the assessment of a given model, as single realizations may vary considerably among themselves ~~(and~~ due to internal variability (and also deviate from the climate evolution observed in reality)~~, such that relying on single realizations possibly deteriorates the assessment of a given model~~. One way to trigger internal variability in GCM simulations is to perturb the initial conditions of the model, leading to several realizations of weather sequences under identical external forcing which also allow to derive a robust distribution of NAO index values. That is why this study is investigating the NAO pattern in a single-model large ensemble (50 members) of a GCM.

However, when interested in NAO impacts on a regional scale, like Central Europe, the GCM is not sufficient for fine-scale responses. Due to their coarse spatial resolution, GCMs are poorly resolving land–water contrasts and topographic ~~properties~~ characteristics which may be highly relevant in climate impact studies over heterogeneous landscapes (Leduc et al., 2019). Thus, dynamical downscaling ~~is advised,~~ of the GCM members using a regional climate model (RCM) ~~(Leduc et al., 2019) . It is however not clear, how global circulation variability affects local climate characteristics when downscaled using an RCM and whether the range of internal variability (i. e., the inter-member spread of the LE) is transferred correctly from the driving GCM into the driven RCM. This question may be important for impact modellers who work with RCM data without taking the driving GCM into account.This study is analysing the NAO in a very large~~ is advised (Leduc et al., 2019) . Such downscaling of a GCM single-model ~~ensemble: the 50 GCM members driving an RCM~~ large ensemble was performed within the Climate Change and Hydrological Extremes ~~(ClimEx) project (Leduc et al., 2019) .It allows for analysing~~ project (ClimEx, www.climex-project.org, Leduc et al., 2019) .

The combination of the driving GCM and nested RCM large ensembles (LE) allows for analyzing the spread of NAO states and responses within one model chain, thus establishing the range of internal variability of the NAO, and finding robust NAO ~~patterns which exceed the uncertainty due to~~ and response patterns by significantly reducing uncertainty associated with internal variability in the ensemble. ~~So this~~ This study also targets the question, how global circulation variability, in this case the NAO teleconnection, affects local climate characteristics when downscaled using an RCM. It specifically aims at evaluating whether the range of internal variability is represented consistently between the driving GCM and the driven RCM. This issue may be important for impact modellers who work with RCM data without taking the driving GCM into account.

To answer these research questions, the study is focussing on four topics and related key questions:

(a) General performance of the model chain: Can the driving GCM resolve the NAO correctly and are climatic implications for Central Europe reproduced?

(b) Nesting approach: Does the RCM correctly incorporate the NAO pattern present in the driving data and produce realistic response patterns?

(c) Internal Variability: What is the range of possible NAO patterns ~~, displayed by the variances~~ and responses, expressed by the inter-member spread (IMS) among the 50 members? ~~Nesting approach: Do GCM NAO impulses propagate correctly into the RCM realizations and produce realistic response patterns?~~

(d) Climate change: How do (a), (b) and (c) change in transient climate simulations until 2099 using an RCP8.5 emissions scenario?

~~This paper is structured as follows: Sect. 2 introduces data and methods and Sects. 3 and 4 present and discuss the results, respectively, followed by conclusions in Sect. 5.~~

2 Data and Methodology

2.1 Data

Data from three different sources ~~are~~ were employed in this study (Table 1). The major source ~~is the ClimEx data set~~ was the LE data set of the ClimEx project which is described in detail in Leduc et al. (2019). The ClimEx project ~~(www.climex-project.org)~~ is conducted in a Québec-Bavarian cooperation and targets issues of hydrological extreme events in the time horizon of 1950–2099, using a nested high-resolution 50 member single-model initial-condition large ensemble with an RCP8.5 emissions scenario from 2006 onwards (Leduc et al., 2019). ~~This single-RCM 50-member ensemble allows for internal variability and extreme events to be detected in high spatial and temporal resolution within a total of 7500 modelled years (Leduc et al., 2019)~~ Five members of the Canadian Earth System Model version 2 (CanESM2 Large Ensemble, 2.8° spatial resolution, Fyfe et al., 2017) with different ocean inital conditions were slightly perturbed in 1950, leading to ten members per ocean family. The members are assumed to become independent about five years after their initialization in 1950 (spin-up-period) (Leduc et al., 2019). Regarding the atmospheric circulation, Fig. 1 shows that owing to the chaotic nature of the atmospheric system the daily NAO index seems to lose dependence from the initial conditions within the course of one month after initialization

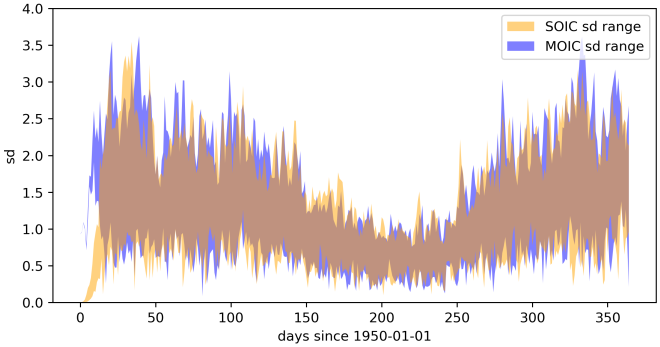

Figure 1. Inter-member standard deviation of a daily NAO index in the CanESM2-LE starting on 1 Jan 1950 as a function of time. The inter-member standard deviation was derived from ten groups of five members with the same ocean initial conditions (SOIC) and ten groups of five members with mixed ocean initial conditions (MOIC, following an approach in Leduc et al., 2019) .

(see Leduc et al., 2019, for a similar presentation of member independence) . As described in Leduc et al. (2019), ~~the original~~ these 50 ~~members of the Canadian Earth System Model version 2 (CanESM2 Large Ensemble, 2.8° spatial resolution, Fyfe et al., 2017) hav been~~ GCM members were dynamically downscaled using the Canadian Regional Climate Model version 5 (CRCM5 Large Ensemble, 0.11° spatial resolution) over two domains ~~in~~ covering Europe and north-eastern North America. During the nesting process, large-scale spectral nudging regarding the horizontal wind field was applied (Leduc et al., 2019). ~~Comparing the single-model~~ This single-RCM 50-member ensemble allows for internal variability and extreme events to be detected in high spatial and temporal resolution within a total of 7500 modelled years (Leduc et al., 2019) . Comparing the internal variability of the CRCM5 members with the ~~inter-model spread of the CORDEX~~ IMS of a subset of the multi-model EURO-CORDEX (Coordinated Regional climate Downscaling Experiment) ensemble regarding winter temperature and precipitation, von Trentini et al. (2019) ~~have shown~~ showed that both ensemble spreads are of comparable magnitude. ~~This similarity~~ The CORDEX ensemble consists of several GCM-RCM combinations set up in a coordinated modelling framework and aims at evaluating uncertainty due to model configuration (Giorgi et al., 2009) . The similarity of the single-model and multi-model spreads suggests that a large fraction of the CORDEX ensemble spread can be explained by internal variability, despite the fact that it was not explicitly sampled within the CORDEX framework ~~(where most models provided a single simulation)~~(where most models provided a single simulation, von Trentini et al., 2019) . Therefore, the GCM and RCM ClimEx ensemble can be expected to capture the ~~most important expressions of natural variability~~ range of winter temperature and precipitation internal variability despite the set up with a single model.

~~The~~ Model data is compared to the ERA-Interim (ERA-I) Reanalysis data set of the European Centre for Medium-Range Weather Forecasts ~~(ECMWF)~~ (Dee et al., 2011, ECMWF) which serves as a reference ~~(REF; Dee et al., 2011) and the~~ (REF). Additionally, a CRCM5 ~~/ERA-I run~~ run driven by ERA-I was used to evaluate the CRCM5 under “perfect” (as far as ERA-I can be assumed ~~as representing~~ to represent reality) lateral boundary conditions (LBC), i.e. without the potential CanESM2

Table 1. Overview of used data sets, their spatial resolution, the number of members and the employed variables.

data name	model type	spatial resolution	members	~~variables~~ model output variable names	institution
ERA-I	re-analysis	$0.75^\circ \times 0.75^\circ$	1	msl [Pa], t2m [K], tp [m]	ECMWF
CRCM5/ERA-I	RCM	$0.11^\circ \times 0.11^\circ$	1	tas [K], pr [$\mathrm{kgm^{-2}s^{-1}}$]	Ouranos
CanESM2	GCM	$2.8^\circ \times 2.8^\circ$	50	psl [Pa], tas [K], pr [$\mathrm{kgm^{-2}s^{-1}}$]	CCCma
CRCM5-LE	RCM	$0.11^\circ \times 0.11^\circ$	50	tas [$^\circ$C], pr [mm]	Ouranos

CCCma – Canadian Centre for Climate Modelling and Analysis

input data error.

The relevant variables for this study ~~are~~were:

- (mean) sea level air pressure (~~msl, psl,~~ referred to as “~~psl~~SLP”, ~~in~~ converted to [hPa]) to obtain the NAO,
- near surface air temperature (~~t2m, tas,~~ referred to as “~~tas~~nSAT”, ~~in~~ converted to [K]),
- total precipitation including liquid and solid precipitation from all types of clouds (~~tp, pr,~~ referred to as “~~pr~~PR”, ~~in~~ converted to [mm]).

ERA-I variables t2m, tp and msl were chosen as they were assumed to most accurately represent the variables from the GCM and RCM models. As the variables derived from the three data sources were originally available with different temporal resolutions (three-hourly for tas in RCM, hourly for pr in RCM, daily in GCM for psl, pr and tas, 6-hourly for ERA-I t2m and msl analysis, and 12-hourly for ERA-I tp forecast data), they were all aggregated to daily values.

Figure A1 shows that the CRCM5 tends to underestimate (overestimate) mean winter nSAT mean in the northern (southern) part of the domain, regardless of the driving data (see first column for ERA-I and third column for CanESM2), whereas winter PR sums are overestimated in nearly the entire domain with strongest values in the south-eastern part. The GCM overestimates (underestimates) nSAT mean north (south) of the Alps. PR sum is underestimated in the entire domain apart from the western side of the Alps in the GCM. However, as this study will focus on changes in nSAT and PR induced by the NAO (see Section 2.2.4), biases are of no large relevance in general, but may show some influence when it comes to regions with particularly high PR sum values.

Commonly, NAO impact studies focus on seasonally aggregated values of the analyzed variables or extreme events (e.g., Stephenson et al., 2006). Yet the NAO, which accounts for variations in the mean zonal atmospheric flow towards Europe, can be assumed ~~not only to~~ to not only influence winter mean values, but also their ~~scattering~~dispersion. So the following analyses ~~are not limited~~ were not confined to winter mean temperature (~~tas~~ nSAT mean) and precipitation sums (~~pr~~ PR sum), selected analyses were also performed on winter mean monthly standard deviations of daily mean temperature (~~tas~~ nSAT std) as a measure of temperature variation.

2.2 Methodology

~~This section gives an overview of the selection of regions and time horizons, the derivation of the employed NAO index as well as corresponding, spatially varying climatic responses in Central Europe and the quantification of internal variability within the study.~~

2.2.1 Regions of interest and time horizon selection

Analyses were performed on time series of spatially averaged information (~~tas mean, pr~~ nSAT mean, PR sum for response variables and ~~psl~~ SLP for index calculation) ~~and~~ as well as on spatially explicit data (~~tas mean, tas std, pr~~ nSAT mean, nSAT std, PR sum). All data were provided as netCDF and most pre-processing was performed using the Climate Data Operators (CDO) of the Max-Planck-Institute for Meteorology (Schulzweida, 2017).

~~Within this study, there are two separated~~ The regions of interest ~~. The first captures the formation region~~ used in this study are displayed in Fig. 2. The formation of the NAO over the North Atlantic ~~using~~ (NAR, AH, IL regions, see annotations in Fig. 2) was analyzed in the ERA-I and CanESM2-LE dataset, while ~~the second one is understood as the response region~~ responses over Central Europe (CEUR, NE, BY, SE) were evaluated in ERA-I, CRCM5/ERA-I, CRCM5-LE and CanESM2-LE.~~Their properties – extent and size in terms of model-specific grid cells – are summarized in Table ??.~~

AH and IL regions are ~~centred~~ centered over Ponta Delgada/Azores and Reykjavik/Iceland, two commonly used stations for NAO index calculations. To avoid micro-climatic impacts and sampling uncertainties of a single gridcell and to account for moving ~~psl centres~~ SLP centres (see e.g., Moore et al., 2013) , both NAO core regions were extended to ~~a~~ 3×3 GCM grid cell ~~matrix each~~matrices. The NAO index proved to be very robust towards the exact shape of the core regions in preliminary analyses.

~~Similar to the pressure centres, the~~ The Central European domain (CEUR) was defined ~~for~~ in the CanESM2-LE by selecting a 5×5 GCM grid cell matrix ~~centred~~ centered over Munich/Germany. This CEUR domain extends from Denmark in the north to mid-Italy in the south and from Poland to France in east–west direction. The corresponding CEUR region ~~in~~ within the ClimEx European domain ~~(Leduc et al., 2019)~~ was used to quantify the impacts of the NAO in the CRCM5-LE dataset. It lies downstream of the westerly flows initiated by the NAO, so the following analyses will set a special focus on the incorporation of large-scale inflow from the western side into the nested RCM.

As the responses to ~~NAO impulses are~~ the NAO were expected to vary over the CEUR domain, it ~~is~~ seemed favourable to analyze spatial structures explicitly in addition to analyses of time series over several subset regions. These subset regions ~~(see e.g. Déqué et al., 2007)~~ (see e.g., Déqué et al., 2007) denote small-scale ~~areas~~sample areas inside the CEUR domain, sized one GCM grid cell each, with expected typical “northern European” (NE) and “southern European” (SE) NAO responses for a more detailed statistical analysis. A third GCM grid cell was chosen to represent the transition zone between NE and SE. Coincidentally, it closely represents the region of Bavaria which is why the name “BY” was assigned to it. ~~The position and size of these subset regions may be derived from Table ?? and Fig. 3.~~ REF and RCM data ~~is spatially averaged over several grid cells.~~ (3×4 and 26×26 grid cells, respectively) was spatially aggregated to GCM resolution for this part of the analysis.

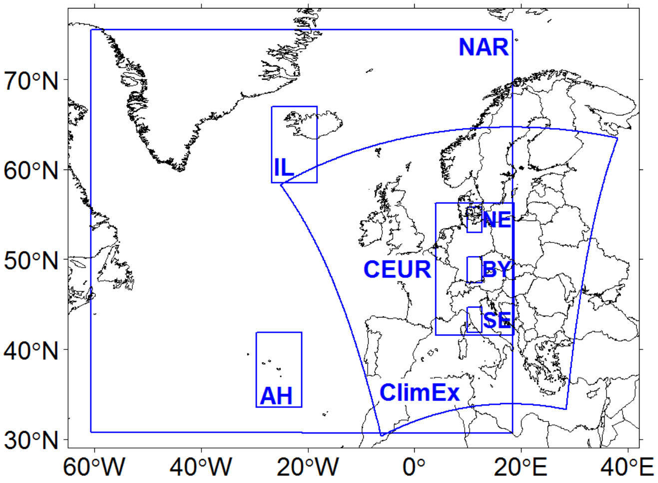

Figure 2. Regions of interest~~: properties~~. ~~Left block: limiting coordinates, right block: count~~ Abbreviations and domain sizes in terms of GCM grid cells ~~within each region. Size of grid cells: see Table 1. Abbreviations~~ (2.8°) are as follows: AH – Azores High (3 × 3); IL – Icelandic Low (3 × 3); NAR – large-scale North Atlantic region ~~for psl composites~~(28 × 16); CEUR – ~~central~~ Central Europe (5 × 5); NE – northern Europe (1); BY – Bavaria (1); SE – southern Europe (1); ~~N/S~~ ClimEx – ~~northern/southern border; E/W – eastern/western border~~domain used in ClimEx project (extent approximately 22 × 12 after resampling to GCM grid).

~~region N E S W ERA-I CanESM2 CRCM5~~

~~AH 42.0°N 21.0°E 33.5°N 29.6°W 12 × 12 3 × 3 – IL 67.1°N 18.1°E 58.6°N 26.8°W 11 × 11 3 × 3 – NAR 75.3°N 18.3°E 30.7°N 60.5°W 105 × 60 28 × 16 – CEUR 55.8°N 18.3°E 41.7°N 4.1°W 19 × 19 5 × 5 129 × 128 NE 55.8°N 12.7°E 53.0°N 9.8°W 3 × 4 1 25 × 26 BY 50.3°N 12.7°E 47.4°N 9.8°W 3 × 4 1 26 × 26 SE 44.7°N 12.7°E 41.7°N 9.8°W 3 × 4 1 26 × 26~~ This study focused on inter-annual analyses which were conducted for two time horizons covering 30 years each. The chosen period length was assumed to include major fluctuations, like ~~several solar cycles or~~ internal climate variations ~~. Thus a representative distribution of NAO events can be expected.~~or several solar cycles, which might affect NAO phases (Andrews et al., 2015) . Thus their influence can be assumed to be represented by the sampled NAO time series.

Relationships between the NAO and ~~weather~~ response variables most probably vary on different time scales (Hurrell and Deser, 2009; Woollings et al., 2015; Xu et al., 2015; Hurrell and Van Loon, 1997). However, as 30 year periods are not long enough for analyses of ~~multidecadal~~ multi-decadal (>30 years) NAO–response variability (Woollings et al., 2015), stationarity in ~~NAO patterns and impacts~~ NAO-impact relationships was assumed for simplicity reasons. The historical (hist; 1981–2010) period was used to establish reference statistics ~~with~~ in the ERA-I data and the ERA-I driven CRCM5 run. These statistics were evaluated in GCM and RCM data to check the models' ability of depicting NAO responses. Links and relationships established for the historical period were also investigated in a far future horizon (fut; 2070–2099).

All data (spatially explicit and subset time series) was aggregated to the seasonal time scale (winter means for ~~tas~~ nSAT and winter sums for ~~pr).~~ PR).

Since the NAO is known to be strongest in winter (Hurrell and Deser, 2009) and the connection between station-based indices and NAO responses tends to be best in winter (see Pokorná and Huth, 2015, for months DJF), analyses were performed for ~~months December, January, February and March (DJFM). First~~ this season only. Preliminary tests had shown that correlations and links between the NAO index and the climate variables were more distinct from noise, if March was included as well. That is why an extended winter season was used here ~~(see also Iles and Hegerl, 2017; Hurrell, 1995)~~ (DJFM, see also Iles and Hegerl, 2017; Hur

2.2.2 Deriving an NAO index

The NAO index was derived from ERA-I and CanESM2-LE data, resulting in 1 REF and 50 ~~model~~ GCM realizations. The ~~index constructed~~ NAO is quantified in this study with an index which is closest to a ~~station-based~~ station based or zonally averaged index. This allowed obtaining an index in a large data set (~~historical and future~~50 members during hist and fut time horizons) at justifiable computational time. ~~It is also easy to interpret in terms of atmospheric physics.~~ Other than indices based on PCA, this index does not represent a “pure” NAO pattern, i.e. the variability of North Atlantic SLP without any other teleconnection patterns like the East Atlantic Pattern (EA) and the Scandinavian Pattern (SCA) (Moore et al., 2013) . Instead, it directly represents the winter SLP gradient over the North Atlantic.

The time series of AH and IL originated from the temporally shortened and spatially averaged ~~psl~~ SLP time series of both grid cell matrices for REF and GCM data only. As the CRCM5 ClimEx domain does not cover the ~~NAR-region (Leduc et al., 2019)~~ AH and IL regions (see Fig. 2), the index ~~can not be~~ was not derived from this data source.

Daily ~~psl~~ SLP values were averaged to monthly means (Cropper et al., 2015) and scaled to obtain ~~average~~ mean $\mu = 0$ and standard deviation $\sigma = 1$, as outlined in Osborn (2004) and Hurrell and Van Loon (1997), by subtracting the 1981–2010 seasonal mean (overbar) and dividing by the 1981–2010 seasonal standard deviation (s_{IL}, s_{AH}):

$$\underline{\mathrm{NAO-Index}}\mathrm{NAOIndex} = \frac{AH - \overline{AH}}{s_{AH}} - \frac{IL - \overline{IL}}{s_{IL}} \tag{1}$$

Monthly indices were next averaged to DJFM means. This approach is similar to Woollings et al. (2015) and Jones et al. (2013). To compare future with historical index values, the future time series of AH and IL were normalized with the present ~~psl~~ SLP standard deviations (see also Ulbrich and Christoph, 1999; Hansen et al., 2017) and mean values. The normalization of each GCM member ~~is~~ was carried out individually, such that each member has specific normalization parameters.

2.2.3 ~~Assessing Climatic Changes Associated with NAO~~Evaluation of the large-scale SLP pattern in RCM data

To estimate whether the NAO may be seen as being correctly represented in the nested RCM data, the reproduction of inter-annual SLP pattern variations in the CRCM5 data was verified. Therefore, monthly mean SLP data of the CRCM5 (both driving data sets) and ERA-I were linearly interpolated to GCM resolution over the ClimEx domain. During interpolation, small scales were automatically filtered such that the remaining large scales of GCM and RCM data may be compared. As a next step, the monthly difference between driving data and the RCM data was taken for each time step and member. From these differences, a root-mean-square difference over the hist and fut time periods was obtained which was later averaged over all ensemble members and the winter months:

$$\mathrm{RMS}^*(i,j) = \left\langle \left\langle \frac{\sqrt{\langle D_m(i,j,t,n)^2 \rangle_t}}{\sqrt{\mathrm{VarDrive}_m(i,j,n)}} \right\rangle_n \right\rangle_{m=12,1-3} \tag{2}$$

$$\mathrm{VarDrive}_m(i,j,n) = \left\langle \left(\mathrm{Drive}_m(i,j,t,n) - \langle \mathrm{Drive}_m(i,j,t,n) \rangle_t \right)^2 \right\rangle_t \tag{3}$$

where $\langle \cdot \rangle$ is the averaging operator over a given index, D_m is the difference between driving data and RCM data; Drive_m is driving SLP data; $\mathrm{VarDrive}_m$ is the variance of the driving data over the 30 year periods; i,j are spatial coordinates of the grid, m are months 12, 1–3, n are ensemble members 1–50 and t are years in 1981–2010 and 2070–2099. The normalization by the square root of the temporal variance of the driving data allows to derive a measure relative to the inter-annual variability of the SLP pattern on a given location. Low RMS^* values indicate a low error.

2.2.4 Climatic Changes Associated with NAO

All data sources (Table 1) were used to obtain response patterns of the ~~given variables~~ variables nSAT and PR. Climatic changes associated with ~~NAO impulses~~ the NAO were evaluated using Pearson correlation coefficients and a slope parameter obtained by linear regression.

ERA-I and CRCM5/ERA-I ~~tas and pr spatial data and subset region time series~~ nSAT and PR data were correlated with the ERA-I index~~time series~~, CanESM2 and CRCM5 members were correlated with the CanESM2 index calculated for the corresponding member.

The correlation analysis assumes (symmetric) linear relationships between the NAO index and ~~tas or pr~~nSAT or PR. So the associated response of the variables to NAO changes ~~can be expressed by the~~ may be quantified by a linear equation (Iles and Hegerl, 2017; Stephenson et al., 2006; Hurrell, 1995):

$$Y = \alpha_1 X + \alpha_0 + \varepsilon_Y \tag{4}$$

with Y being the (response) variable at a given grid cell that is partly explained by the NAO (X, the predictor) and by any other influences (ε_Y; Stephenson et al., 2006; von Storch and Zwiers, 2003). The coefficient α_1 was estimated on each grid cell using ordinary least squares regression with the R function lm ~~. It represents the average change in tas or pr that accompanies one~~ (www.rdocumentation.org). It represents mean change in nSAT or PR that accompanies unit index change during the time period under consideration (Iles and Hegerl, 2017). The line offset α_0 in Eq. (~~2~~4) equals the long-term mean. The α_1 coefficients may be computed with respect to normalized index series (von Storch and Zwiers, 2003), but in this study the non-normalized index time series was preferred in order to take into account the member-specific index units.

2.2.5 Addressing Internal Variability

~~The~~ Internal variability was understood as being represented by the oscillations around the long-term mean of the time series of a given variable (Hawkins and Sutton, 2011) . In this point of view, IMS of the LE originates from the superposition of all 50 realizations with their respective inter-annual variability. As the climatic evolution of all 50 members is equally likely by construction of the ensemble, this spread represents an envelope of possible sequences of weather events at any given time step or location. This allows to sample internal variability at single points in time as the range of the members' values.

Therefore, the NAO–response relationship was analyzed individually for each GCM and RCM member (as is done e.g. in Woollings et al., 2015). ~~Ensemble averages partly mask internal model noise (Zwiers and von Storch, 2004) , but also the spread of internal variability which this study is addressing. Nevertheless,~~

Aggregations to ensemble means (like in Deser et al., 2017) and standard deviations (sd, see also Leduc et al., 2019; Déqué et al., 2007) , the latter representing the IMS in maps, were only performed for illustrating purposes ~~the results were aggregated to ensemble averages (like in Deser et al., 2017) . In~~ in order to avoid ~~suppressing ensemble scattering in the spatial approach, the inter-member spread is represented in terms of the standard deviation of all 50 members on a given grid cell (std.dev50, see also Leduc et al., 2019; Déqué~~

masking model internal variability (Zwiers and von Storch, 2004) .

3 Results

The result section is structured in two large parts: Section 3.1 deals with the representation of the NAO and climatic responses in the GCM and RCM and Section 3.2 targets internal variability in the GCM and RCM. Section 4 will follow the structure as defined by the four key questions.

3.1 NAO within the ClimEx Data Set

3.1.1 NAO index and ~~psl~~ SLP conditions

First, a ~~REF~~ reference NAO index was calculated from the ERA-I reanalysis. It is found to be in good accordance with often cited NAO indices like the time series of Hurrell (Pearson correlation of $r = 0.95$ with REF NAO index; index available at https://climatedataguide.ucar.edu/climate-data/hurrellnorth-atlantic-oscillation-nao-index-station-based)~~and therefore serves~~ . For further analyses it will therefore serve as a reference.

The CanESM2-LE produces NAO index values which follow a distribution similar to the ERA-I data (~~centred~~ centered over zero, slight ~~left-skewness), though the~~ surplus of low positive NAO values, see Fig. 3 (a)). The CanESM2-LE distribution appears smoother due to a larger sample size (~~see Fig. 2 (a))~~$n = 1500$ for CanESM2-LE and $n = 30$ for ERA-I). Maximum and minimum ~~values~~ index values (x-axis in Fig. 3 (a)) of some of the 50 members exceed those of the REF realization; thus, the REF realization ~~comfortably lies~~ lies well within the ensemble ~~spread~~IMS. The future NAO index shows a similar distribution of values, but with slightly less positive and more negative values ~~.~~ (red curve in Fig. 3 (a)).

~~Pairwise correlationsbetween the members and~~ For further analyses on the IMS as a measure of internal variability, the

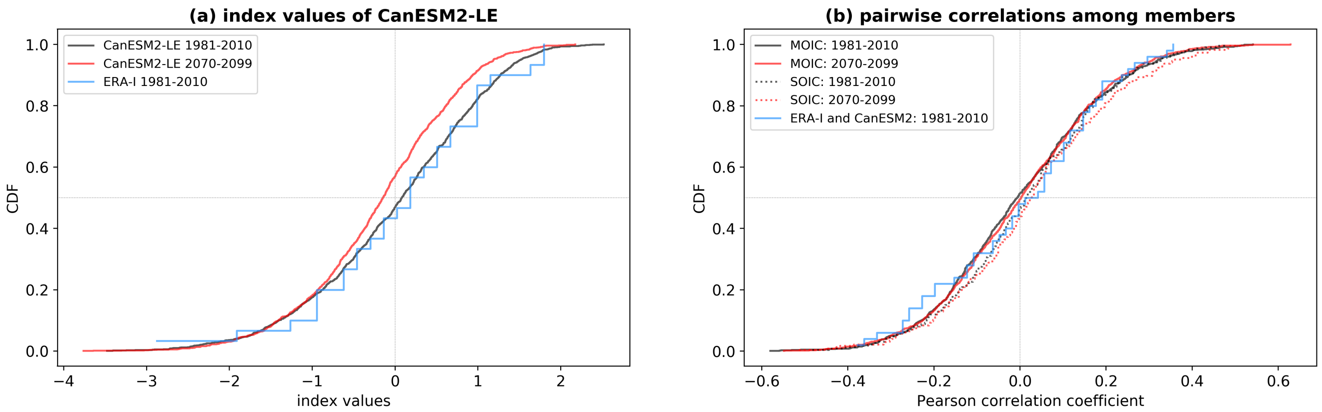

Figure 3. ~~Comparison~~ Cumulative density functions (CDFs) of ~~CanESM2-LE and ERA-I~~ NAO index values. ~~Left:~~ (a) distribution of all CanESM2-LE (~~grey for 1981–2010 and red for 2070–2099~~$n = 50 \times 30$ per period) and ERA-I (~~blue~~$n = 30$) NAO index values. ~~Right: distribution of~~ (b) pairwise correlations ~~between all 50 CanESM2-LE~~ among member NAO index time series from the same ocean families (~~grey for 1981–2010 and red for 2070–2099~~SOIC – same ocean initial conditions, dotted lines, $n = 225$), from different ocean families (MOIC – mixed ocean initial conditions, solid lines, $n = 1000$) and ~~CanESM2-LE members with~~ between ERA-I and all CanESM2 members (~~blue~~$n = 50$). Black: 1981–2010 CanESM2-LE, red: 2070–2099 CanESM2-LE, blue: 1981–2010 ERA-I.

independence of the 50 ensemble members is of high importance. To investigate independence among the ensemble members in both 30 year time frames, it seems favourable to analyse pairwise member correlations. Although zero correlations do not automatically imply independence, clear correlations among members would contradict the assumption of independence. In order to take into account the two perturbations during the production of the LE (1850 for 5 ocean families, 1950 for perturbations leading to 10 members per ocean family), these correlations were split in two groups like in Leduc et al. (2019) : (i) correlations among the 10 members from the same ocean family (same ocean initial conditions in 1950, SOIC, $n = 225$, see dotted lines in Fig. 3(b)) and (ii) correlations between each member and the ~~ERA-I time series in general are not strong as can be seen in Fig. 2~~ 40 members from the 4 other ocean families (mixed ocean initial conditions, MOIC, $n = 1000$, see solid lines in Fig. 3 (b)~~, highlighting the independence of~~).

These correlations approximately follow a normal distribution with $\mu = 0$. There is a slight surmount of low positive correlations in the SOIC group compared to the MOIC group which is (not significantly) stronger in the fut time horizon (see red and black dotted lines in Fig. 3 (b)). In general, the members are thus not seen as being dependent. As will be discussed below, the SLP pattern over the North Atlantic changes slightly in the future period. So the direct comparison between historical and future SOIC and MOIC correlations remains difficult. The members also show no systematic correlation with the REF NAO index despite similar statistics (see also Fig. 9). Thus, the ~~CanESM2-LE members in terms of internal variability (see also Fig. A2). It is also visible that correlations among members are not stronger or weaker than correlations of members and ERA-I. They are not systematically related to the~~ ERA-I ~~(the “reference”) realization~~and GCM indices can be seen as not dependent realizations

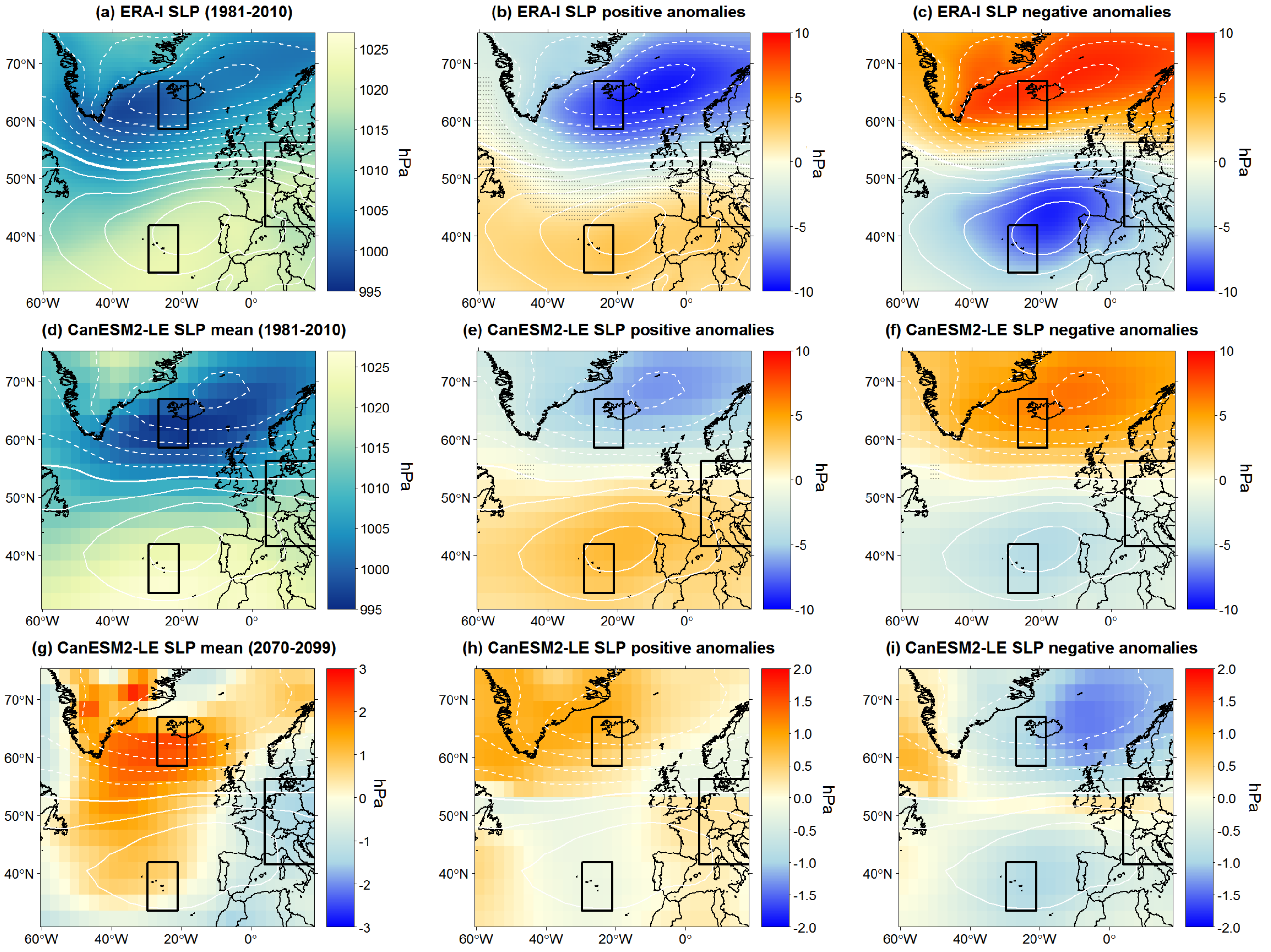

Figure 4. ~~Winter~~ NAR winter mean ~~sea level pressure (MSLP)~~ SLP [hPa] ~~for~~ composites in REF ((a)–(c)) and GCM ((d)–(~~f~~i)) ~~composites (1981–2010)~~ data showing long-term ~~average~~ neutral conditions (left column), NAO positive (mid column) and negative anomalies (right column)~~over the North Atlantic region~~. (a)–(f): for 1981–2010. (g)–(i): 2070–2099 changes ~~towards~~ with respect to 1981–2010 in GCM data. White isolines: difference between positive and negative ~~(dashed)~~ anomalies by a step of 2.50 hPa, as e.g. in Hurrell (1995), solid: positive, dashed: negative, bold line: zero. ~~Black boxes~~Stippling in subpanels (a)–(f): regions ~~used for index calculation over~~ where the ~~North Atlantic and response region in central Europe~~anomaly is smaller than the standard error of the composite samples. ~~Red~~ Black boxes: ~~position of subset~~ AH, IL and CEUR regions ~~in~~ (see Fig. ~~8~~2).

drawn from the same distribution.

~~Typical spatial psl features~~ In addition to the NAO index evaluation, Fig. 4 presents the large-scale SLP patterns in the NAR region ~~are shown in Fig. 3 for average~~during neutral, positive and negative NAO conditions. Positive ~~and negative~~(negative) index years are chosen, if the respective ~~absolute~~ index value exceeds 1 (−1) as in Rogers (1984). ~~This stratifies the original~~

~~data in three subsets with positive, negative and indifferent index values. Under average psl~~ The neutral conditions refer to the 30 year SLP average. Under neutral NAO conditions, the North Atlantic region is characterized by a pressure dipole. This structure is intensified and tilted clockwise in the CanESM2-LE ensemble mean (middle row of Fig. 4) compared to REF (top row). The mean SLP difference between the CanESM2-LE mean and REF reaches up to 10 hPa in both directions~~, with the strongest overestimation above Greenland and underestimation~~. SLP values are higher over Greenland and lower over the North Sea ~~. The~~ in the CanESM2-LE compared to ERA-I (panels (a), (d) in Fig. 4). Long-term neutral states of both data sources show robust signals in the entire NAR region (i.e. no stippling). This suggests that the different patterns in GCM and REF data are not singularly artefacts arising from different sample sizes. The GCM multi-member composites of positive and negative phases ~~also~~ show less pronounced ~~psl~~ SLP anomalies than the REF data. ~~The difference between psl~~ This difference between GCM and REF may be due to the fact that REF composites were derived from $n = 3$ negative and $n = 4$ positive years whereas the GCM data provided $n = 264$ negative and $n = 263$ positive years during 1981–2010. Regions with strong sampling uncertainties, i.e. where the standard error is larger than the anomaly, are indicated with stippling in panels (a)–(f). These regions are mostly found in the transition region between the wider ERA-I AH and IL nodes, whereas the SLP anomalies at the NAO centres of action show less uncertainty. The GCM patterns are more robustly assessed (i.e. less prone to sampling uncertainty) as can be seen by the very small area with stippling in which the sign of the anomaly may not be assessed robustly. The difference between SLP anomalies in positive and negative years representing the pressure variability is indicated by white lines. ~~It~~ This difference is weaker in the CanESM2-LE mean than in ERA-I data, but located in similar regions. These NAO centres of action reach GCM (REF) ~~psl~~ SLP differences between positive and negative conditions of about 12.5 (17.5) hPa in the IL region and 7.5 (10.0) hPa in the AH region. They do not coincide with the highest and lowest ~~average psl values~~ SLP values in the neutral state, but are situated near the 3×3 ~~pixel grid~~ GCM grid cell matrices used for index calculation. This is very promising as it ~~backs~~ supports the choice of these ~~psl~~ SLP centres for index calculation.

Under projected future climate conditions, ~~psl~~ SLP rises over large parts of the North Atlantic and shows less variability (see Fig. ~~3~~ 4 (g)–(i)). Future positive phases tend to be weaker as ~~psl~~ SLP shows a marked increase in the northern NAO node region. Negative phases exhibit ~~psl~~ SLP decreases in both node regions, although with larger changes ~~in the northern region~~near IL, resulting in negative phases to become slightly weaker as well.

3.1.2 ~~Local climate response to the NAO~~

Having established a reasonably plausible representation of the NAO in the driving data, the next step is to evaluate the large-scale NAO pattern in the RCM data. Figure 5 maps the RMS* of the difference between driving data and RCM SLP during 1981–2010 for driving data ERA-I (a) and CanESM2 (b) and 2070–2099 (c). A value of RMS$^* \geq 1$ indicates that the root-mean squared error between the RCM and driving data is larger than the temporal variability in the driving data. In this case, the large-scale SLP pattern may not be seen as being correctly represented in the RCM data. The large-scale SLP pattern over the entire ClimEx domain, which also includes the CEUR, NE, BY and SE domains, is reasonably well represented: with RMS$^* < 1$ in most parts of the entire ClimEx domain for both driving data sets and both periods

~~Spatial patterns of change in tas mean (in) for a unit change in the NAO index for ERA-I, CRCM5/ERA-I, CanESM2-LE and CRCM5-LE in 1981–2010 ((a)–(f)) and the difference to 2070–2099 ((g)–(j)). Both 50-member ensembles are represented with ensemble mean and std.dev50. Blue isolines show the respective correlations by an increment of 0.1 (thick line is zero-correlation, solid lines positive).~~

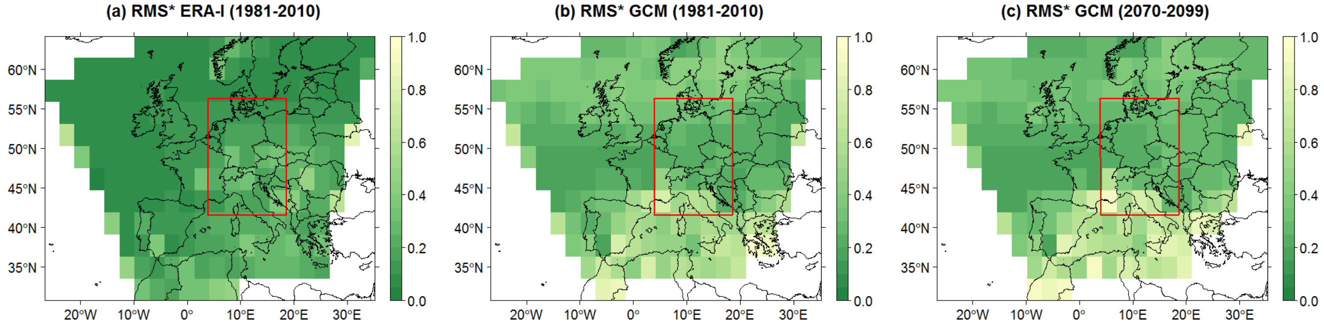

Figure 5. ~~Spatial patterns~~ RMS* of ~~distributed change in pr sum~~ monthly SLP differences between driving data and CRCM5 members, calculated following Eq. (~~in~~ 2)~~for~~. Colouring: RMS$^* \leq 1$ significant at $p \leq 0.05$ with a ~~unit index change in the NAO index~~ false detection rate smaller than 0.1 (see Wilks, 2016) . (a) for driving data set ERA-I ~~, CRCM5/ERA-I, CanESM2-LE and CRCM5-LE in~~ (1981–2010), (~~(a~~b) ~~–~~for driving data set CanESM2 (~~f~~1981–2010)~~) and the difference to 2070–2099~~ , (~~(g~~c) ~~–~~for driving data set CanESM2 (~~j~~2070–2099)~~)~~. ~~Both 50-member ensembles are represented with ensemble mean and std.dev50. Blue isolines show the respective correlations by an increment~~ Red box: position of ~~0.1 (thick line is zero-correlation, solid lines positive, dashed lines negative correlations)~~CEUR domain.~~Note that the colour bar in the bottom row is flipped compared to Fig. 4.~~

(significant at $p \leq 0.05$ using a t-test with a false detection rate < 0.1 to account for multiple hypothesis testing, see Wilks, 2016) . All data sets show an RMS* increase towards the south, indicating that in these regions the control exerted by the lateral boundary conditions on the CRCM5 internal solution appears to be weaker. The RMS* is larger in the CanESM2/CRCM5 combination than in the ERA-I/CRCM5 combination, and slightly increases in the future period in the southern parts (see Fig. 5 (c)). In the CEUR domain (indicated as red box in Fig. 5), however, errors are low in general and therefore the NAO pattern of the driving data may be assumed to be correctly incorporated there. It is thus reasonable to continue with the evaluation of nSAT and PR responses in the CEUR domain.

3.1.2 Local climate response to the NAO

~~NAO tas and pr~~

NAO nSAT and PR spatial responses as revealed in the ERA-I data are reproduced in their general properties under current climate conditions in the CanESM2-LE and CRCM5-LE (see Figs. ~~4, 5, 6). Strongest responses~~ 6–8). Responses with the highest α_1 magnitudes occur in the CRCM5/ERA-I run for all variables. Regarding the absolute α_1 values, the CRCM5-LE values meet the ERA-I values better than the CRCM5/ERA-I run.

Positive NAO conditions are accompanied by winters with warmer temperatures (up to +2 K per unit index change, see Fig. ~~4~~6) and less day-to-day ~~tas~~ nSAT variability (see Fig. ~~6).~~ 7). The generally positive relationship between ~~tas~~ nSAT mean and

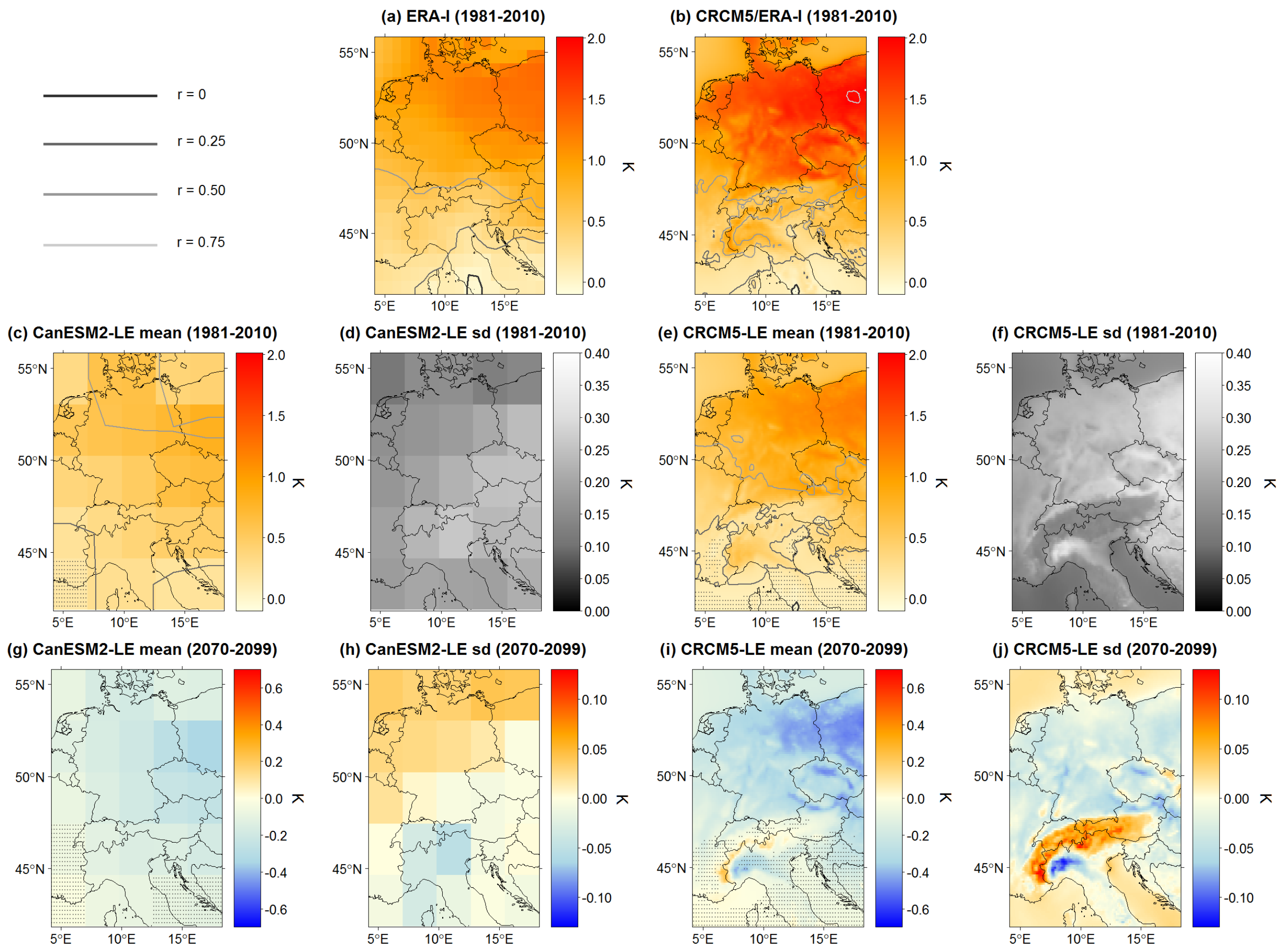

Figure 6. Spatial patterns of ~~distributed~~ change in ~~tas std~~ nSAT mean (in [K]) for a unit ~~index~~ change in the NAO index for ERA-I, CRCM5/ERA-I, CanESM2-LE and CRCM5-LE in 1981–2010 ((a)–(f)) and the difference ~~to~~ of 2070–2099 with respect to 1981–2010 ((g)–(j)). Both 50-member ensembles are represented with ensemble mean and ~~std~~sd representing the IMS. ~~dev50. Blue isolines show~~ Grey lines in the ~~respective correlations by~~ ensemble mean maps represent the Pearson correlation between nSAT mean and the NAO index at an increment of ~~0.1 (thick line is zero-correlation, solid lines positive, dashed lines negative correlations)~~0.25; grey shadings see legend in upper left panel. ~~Note that~~ Grey stippling in the ~~difference maps for CanESM2-LE and CRCM5-LE~~ ensemble mean maps show regions were ~~calculated using absolute values~~SNR < 1, SNR being the signal-to-noise ratio between the 30 year ensemble mean and sd of GCM and RCM LEs in both time periods.

NAO (see Fig. ~~4~~6) is strongest in the ~~eastern~~ north-eastern parts of the domain. Regionally, the NAO explains up to 40–60 % of ~~tas~~ nSAT mean variability (see also Fig. ~~?? where the tas~~ A2 where the nSAT mean α_1 share of the entire winter standard deviation of daily temperature values is shown). Explained variance is highest in the CRCM5/ERA-I run and lowest in the CanESM2-LE. The reduction of ~~tas~~ nSAT variability reaches up to 0.4–0.6 K in the ~~northern~~ northeastern continental section

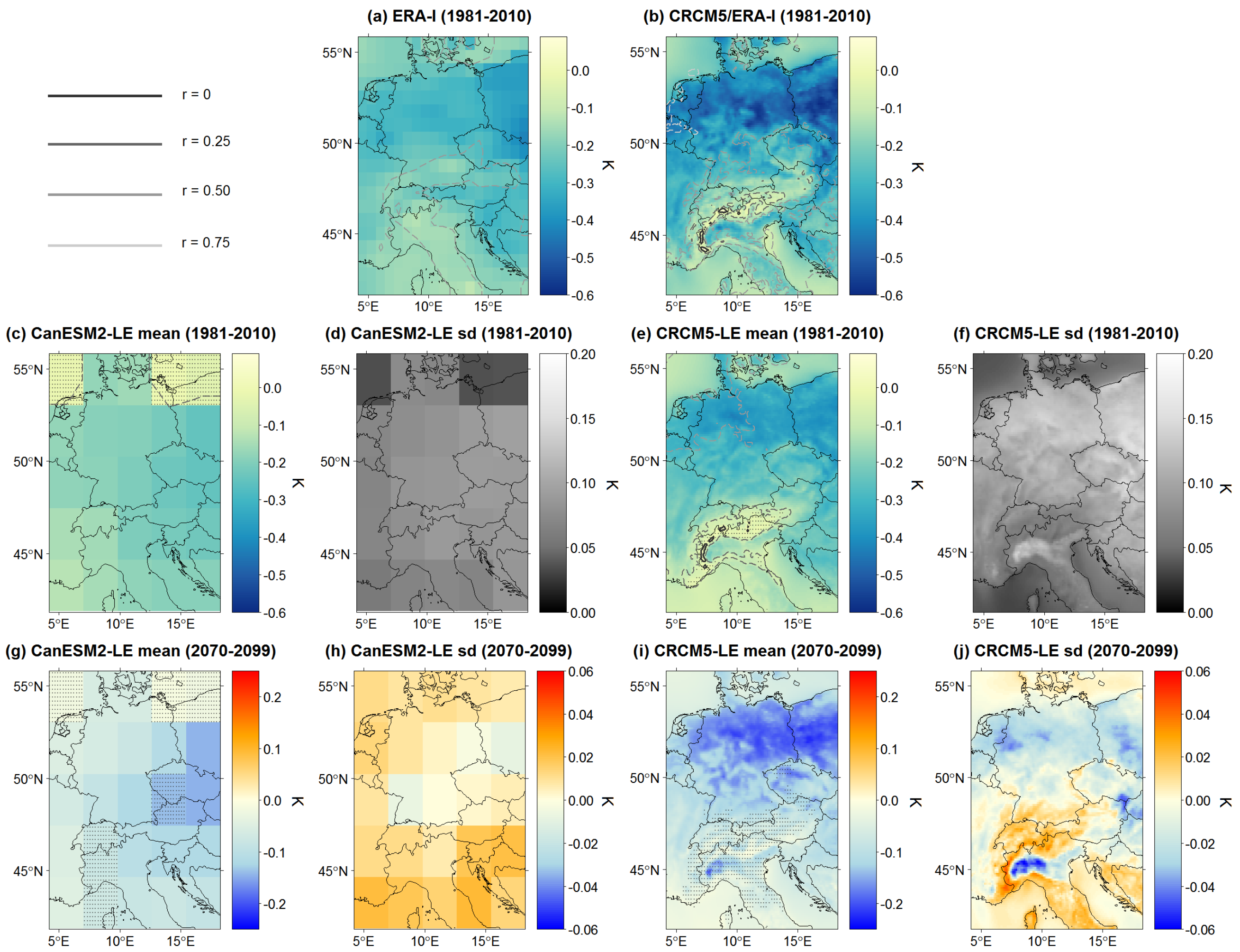

Figure 7. Like Fig. 6, but for nSAT std (in [K]). Note that the difference maps for CanESM2-LE and CRCM5-LE mean were calculated using absolute values.

while it is near zero in the southern part of the domain. ~~In Fig. 6, the spatial patterns of ERA-I and CRCM5/ERA-I differ stronger than in Fig. 4, but CRCM5/ERA-I and LE mean agree well.~~

~~Positive~~ In comparison to the neutral state, positive phases are also accompanied by more humid conditions in the north~~(because of more pr)~~, and drier conditions in the south of the CEUR domain (~~because of less pr,~~ see Fig. ~~5~~8). The strength of the ~~NAO–pr~~ NAO–PR relationship, r, is not affected by topography in any of the models within the domain; only the pivotal line crossing Europe is following the Alpine ridges ~~.~~ (see solid dark line in Fig. 8, panels (a)–(c) and (e)). The change between positive and negative r and α_1 occurs within a very narrow region. Within the CanESM2-LE, this zero-line is shifted northwards compared to ERA-I, CRCM5/ERA-I and CRCM5-LE. As is visible in Fig. ~~5~~8, higher α_1 values in mountainous regions indicate strong

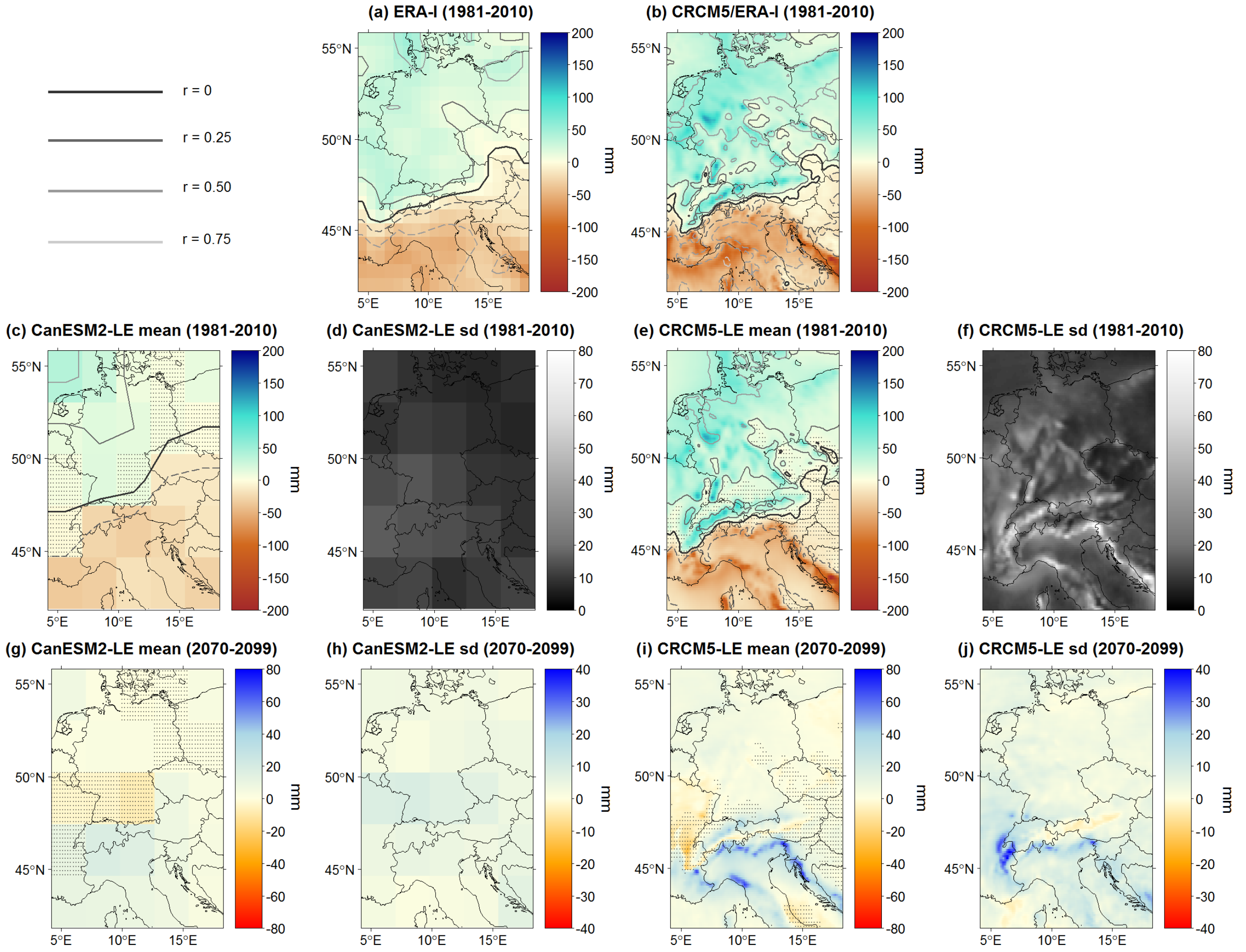

Figure 8. Like Figs. 6–7, but for PR sum (in [mm]). Note that the difference maps for CanESM2-LE and CRCM5-LE mean were calculated using absolute values and that the colour bar in the bottom row is flipped compared to Figs. 6–7.

~~responses to NAO with~~ NAO responses related to orography. Regionally, the NAO accounts for 40–50 % of total ~~pr~~ PR sum variance, in both positively and negatively correlated regions. In the CRCM5-LE, single spots in mountainous regions (e.g., in the Dinaric Alps) show extremely high ~~pr~~ PR sum α_1 values (up to ± 220 mm per unit index change) ~~, stressing~~ where long-term mean PR sums are also very high. This stresses the more detailed production of geographical features~~in the .~~, but also the tendency to evolve local extreme values in the high-resolution RCM (see similar results for local daily extreme precipitation in Leduc et al., 2019) ~~. Pr~~ which may even be noted in the (spatially aggregated) bias towards the GCM (see Fig. A1 (f)). PR sum shows only weak correlations ~~and changes~~ in the central region of the CEUR domain.

The mean state of ~~tas and pr~~ nSAT and PR changes in the transient climate simulation towards warmer and moister conditions

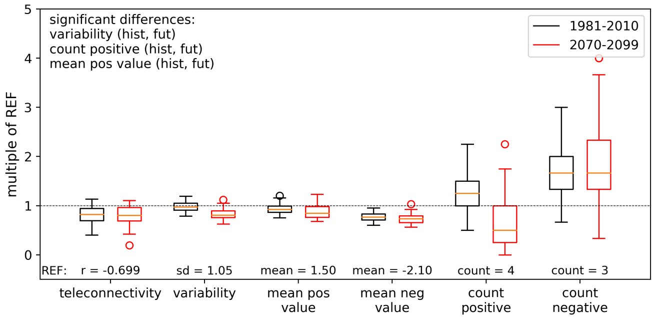

Figure 9. Several index statistics of all 50 CanESM2-LE members expressed as multiples of the respective ERA-I value (REF value set to 1.0): teleconnectivity (Pearson correlation between AH and IL time series), index variability (expressed as ~~sd~~ standard deviation in time of index time series), mean value of all positive (negative) phases ~~,~~ and count of all positive (negative) phases in a single realization. Positive (negative) years are defined by an absolute index value exceeding 1. Text in upper left corner: significantly ($p \leq 0.05$, using an unpaired Mann-Whitney/U-test) different outcomes in the ~~future~~fut time frame.

with less intra-seasonal variability of ~~tas~~nSAT. For a detailed description of the future climate evolution (though for 2080–2099) in Europe within the CRCM5-LE see Leduc et al. (2019). Future ~~correlations~~ NAO–climate relationships weaken in general compared to the historical ones ~~(apart from some regions in the south with increasing r for pr) for all variableswhich adds more noise to the signal: the signal-to-noise ratio (SNR) between LE ensemble means and std. dev50 is reduced for both CRCM5-LE and CanESM2-LE for tas and pr in the future period (see Fig. **??** and A4). This finding is valid for r and α_1. Less variance of tas and pr is explained by the NAO (see α_1 distribution in Fig. 4 and Fig. 5, as well as Fig. **??** for tas mean). Also, tas std decreases less with unit index change in the projected future climate (Fig. 6 (g), (i), blue shading).~~ for all variables. The spatial patterns of NAO-induced change ~~though~~ do not change considerably between both periods. The response to ~~NAO impulses~~the NAO, α_1, is clearly reduced in ~~tas~~ nSAT mean as is ~~tas~~ nSAT std, and there is also a reduction in ~~pr sum change .~~ PR sum change (panels (g), (i) in Figs. 6–8).

3.2 ~~Transferring~~ Internal Variability ~~from~~ at the GCM ~~to the~~ and RCM scale

The ~~transfer~~ representation of internal variability ~~from the GCM to the RCM~~ in the GCM and RCM regarding the responses to the NAO in CEUR and subset regions NE, BY, SE is assessed via ~~the difference in the inter-member spread~~ differences in the IMS of the CRCM5-LE compared to the CanESM2-LE.

3.2.1 Multi-member ensemble

The CanESM2-LE reproduces typical index characteristics: Fig. ~~7~~ 9 summarizes several statistics for all 50 GCM members as multiples of the REF value. Generally, the ensemble meets the REF value in all aspects of the NAO index. However, some GCM members only reach half of the REF teleconnectivity values (minimum: $r = -0.281$, not ~~significant~~ significantly different from zero at $p \leq 0.05$ using a t-test; REF $r = -0.699$). This finding is especially interesting as this metric quantifies the strength of the NAO within the individual members. ~~Internal variability of index characteristics is prone to significant changes between present and future conditions as is also seen in Fig. 7. The ensemble spread~~ The IMS of the teleconnection strength, though, does not change significantly over time, in spite of the ~~psl~~ SLP changes over the North Atlantic. The 2070–2099 NAO index exhibits less inter-annual variability, less positive phases, more ~~indifferent phases and thus~~ neutral phases and a relative increase of negative phases but with reduced mean values (see also Fig. ~~2~~ 3 (a)).

The spatial expression of NAO response internal variability in the form of diverging ensemble members can be derived from Figs. ~~4, 5, 6~~6–8 (subplots (d), (f)) presenting spatially distributed ~~std. dev50.~~ ensemble sd as a measure for IMS. Largest deviations for ~~tas~~ nSAT mean are found in continental regions of CEUR, but they do not ~~generally~~ simply correspond to high or low α_1 ~~. Low std. dev50 corresponds~~ (see also Fig. A3 (a)–(d)). Low IMS corresponds mostly to Alpine and sea regions. For ~~tas~~ nSAT mean, the ~~SNRs~~ signal-to-noise ratios (SNR) between ensemble mean and ~~inter-member spread~~ sd exceed 1 in most regions north of the Alps (see ~~Fig. ??). Regarding pr~~ regions without stippling in Fig. 6). nSAT std shows SNR<1 in the northern parts of the CanESM2-LE data (see Fig. 7 (c)) and in the Alpine region of the CRCM5-LE data (Fig. 7 (e)). This variable shows a strong linear relationship between LE mean and sd (Fig. A3 (e)–(h)). Regarding PR sum, RCM members vary most in regions with highest absolute α_1 values and altitudes, ~~while high α_1 of tas std are accompanied with large inter-member spreads for GCM and RCM. For pr~~ but there is no clear dependence in GCM (Fig. A3 (i)–(l)). For PR sum, there is an east-west corridor of SNR values below 1 which accompanies rather low ~~correlation~~ α_1 values (see Fig. ~~A4). This indicates that low ensemble average correlations in this region are mostly due to diverging correlation values among the single members (noise). The SNR shows similar spatial distributions in r and α_1 for tas mean and pr sum.~~ 8). In addition to future changes in the NAO responses ensemble means, there is also a change in the spatial distribution of the ~~std.dev50 values (see Figs. 4, 5, 6,~~ IMS expressed as ensemble sd (see subpanels (h), (~~i)~~ j) in Figs. 6–8).

~~Figure 8 illustrates the inter-member spread in the spatially averaged subset regions for~~ To further investigate the IMS, Fig. 10 illustrates the Pearson correlation coefficients r between the NAO index and ~~tas mean or pr sum . Both ensembles~~ subset regions nSAT mean or PR sum in GCM and RCM LEs separately. Both ensemble IMS generally envelope the REF value (dashed line) of the given region, apart from GCM hist in ~~tas BY.~~ Fig. 10 (b). This finding does not change in the projected future climate: most boxes and whiskers keep their size, though ~~pr SE (GCM)~~ GCM PR SE is characterized by a smaller range in the future and GCM nSAT NE by a larger one (significant at $p \leq 0.10$ and $p \leq 0.05$, respectively, using an F-test for comparison of variances)~~and tas NE (GCM) by a larger one ($p \leq 0.05$, F-test). The (sometimes significantat $p \leq 0.05$, using an unpaired Mann-Whitney/U-test) shift of boxes~~. Ensemble mean values exhibit a (sometimes significant) shift towards lower r values in the future for both models ~~is clearly visible for tas mean and pr sum.~~ in some regions for nSAT mean and PR sum.

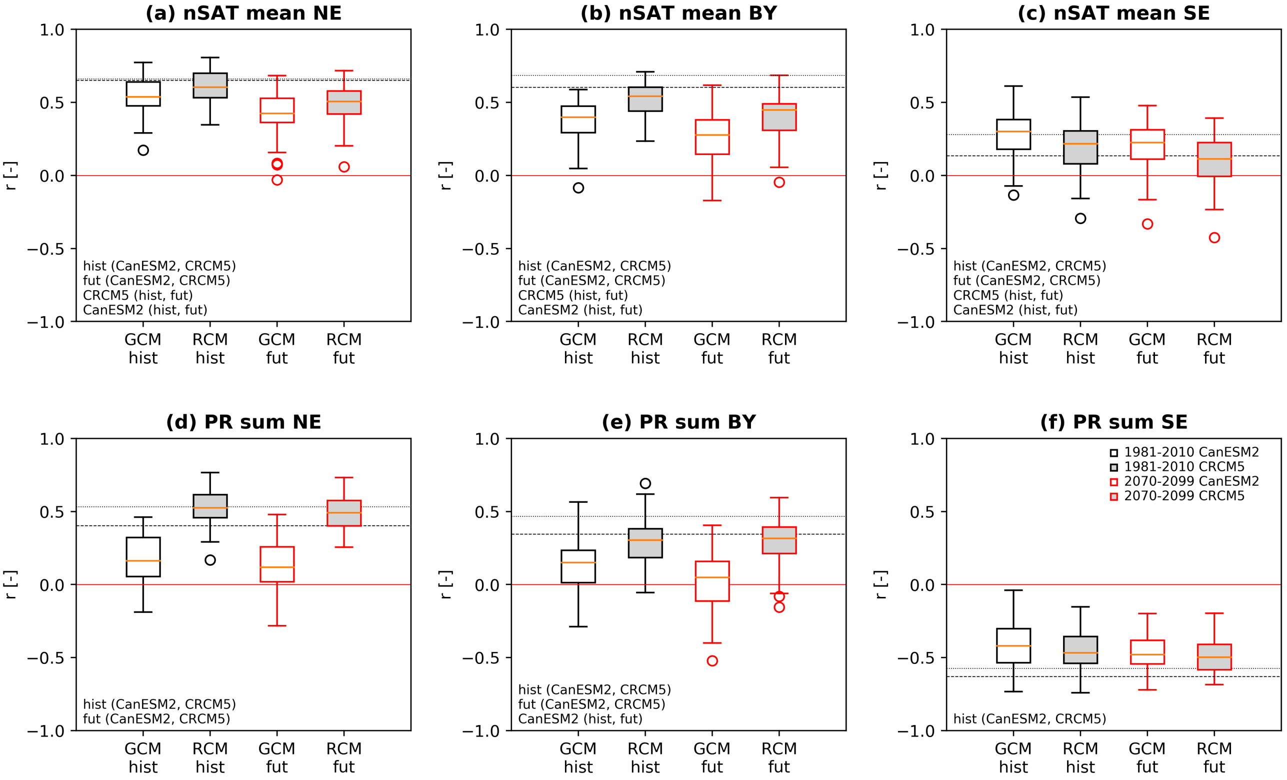

Figure 10. Boxplots of ~~tas~~ nSAT mean ((a)–(c)) and ~~pr~~ PR sum ((d)–(f)) showing ~~r~~ Pearson correlation of 50 CanESM2-LE and CRCM5-LE (grey filling) realizations for three regions (NE, BY, SE) in historical (black outlines) and future (red outlines) time horizons. Dashed (dotted) horizontal lines indicate the ERA-I ~~REF~~ (CRCM5/ERA-I) value; text denotes combinations of which the differences are significant ~~with~~ at $p \leq 0.05$ using an unpaired Mann-Whitney/U-test for the comparison between hist and fut periods and a paired Wilcoxon test for the comparison between GCM and RCM LE.

An unpaired Mann-Whitney/U-test was applied here as the samples from hist and fut were seen as being drawn from different climates (using a χ^2-test, the null hypothesis of independence between hist and fut periods could not be rejected at $p \leq 0.05$).

3.2.2 Change of scales

Having analyzed GCM and RCM separately so far, it is now advised to compare both ensembles. A χ^2-test revealed that GCM and RCM samples of r can be seen as significantly dependent in both time frames. The amount of variance explained by the NAO is generally higher in REF than in the RCM ensemble mean, which in turn is higher than the GCM ensemble mean. The CRCM5-LE enhances the relationship showing higher r and α_1 values than the CanESM2-LE (see Fig. ~~8~~ 10 for r, where hist(CanESM2, CRCM5) or fut(CanESM2, CRCM5) is indicated; but also Figs. ~~4, 5 and 6~~6–8 for α_1). This enhancement by the CRCM5 is ~~notable~~ notably independent of the driving data: for both variables, the CRCM5/ERA-I r value (dotted lines in

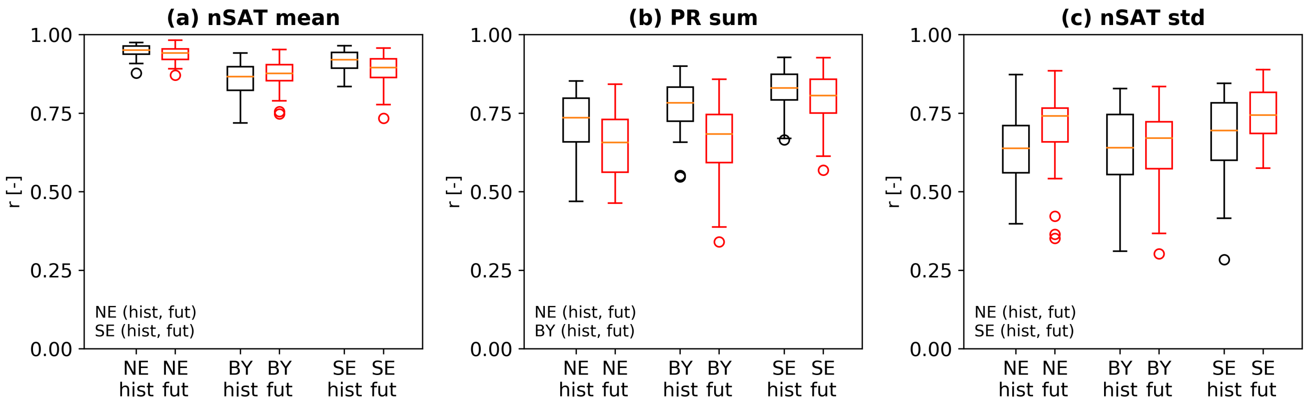

Figure 11. ~~Similarity~~ Temporal co-variability of ~~matching~~ CanESM2 and CRCM5 subset ~~region (NE, BY, SE)~~ regions in all 50 members. Each boxplot represents 50 Pearson correlation coefficients between the time series ~~(tas~~ of variables nSAT mean (a), ~~pr~~ PR sum (b) ~~expressed as boxplots~~ and nSAT std (c) in the subset regions of ~~r between matching GCM~~ CanESM2 members and ~~RCM~~ the corresponding CRCM5 members. ~~Hist~~Time periods used for correlations: 1981–2010 (hist, ~~fut:~~ black), 2070–2099 (fut, red). For regions NE, BY, SE see Fig. 2. Text denotes combinations of which the differences are significant at $p \leq 0.05$ using an unpaired Mann-Whitney/U-test.

Fig. ~~8~~10) is also found to be higher than the ERA-I value in most regions (dashed lines in Fig. ~~8; see also Fig. A3, first column, for spatial distribution of the CRCM5/ERA-I error).~~ 10). In all subset regions, the CRCM5/ERA-I r value lies in the upper part (stronger correlations) of the CRCM5-LE ensemble values.

Figure ~~8~~ 10 shows that mean r values of RCM (grey filling) and GCM (hollow) members are significantly different in all subset regions for ~~tas~~ nSAT mean in both time horizons, but only in the NE and BY regions for ~~pr~~ PR sum; in SE, ~~effectively no difference~~ only weak differences between GCM and RCM ~~pr~~ PR sum r distributions ~~is visible. Apart from pr NE (both time horizons) no significant (at $p \leq 0.05$, F-test) box size (spread amplitude) difference between GCM and RCM is visible.~~ are visible. In NE and BY this difference is expressed by higher r values in RCM data, whereas in the SE region lower r values are found in the RCM data (only for ~~tas~~ nSAT mean). ~~Thus the inter-member spread of~~ Apart from PR sum in the NE region (both time horizons), no significant difference between the spread amplitudes of GCM and RCM is visible ($p \leq 0.05$, F-test). The IMS of the correlation between NAO and response variables is not generally altered during the nesting process.

~~When correlating matching subset region time series (see Fyfe et al., 2017, for a similar approach) of CanESM2-LE~~ To evaluate the co-variability of CanESM2 and ~~CRCM5-LE as~~ CRCM5 data in the subset regions, time series of the response variables originating from both data sources were correlated member-wise (see Fyfe et al., 2017, for a similar approach) . As can be seen in Fig. ~~9~~11, highest accordance on average is reached for ~~tas mean~~ nSAT mean in both periods, indicating that CanESM2-LE and CRCM5-LE show very similar temporal variability for this variable. The co-variability of GCM and RCM time series is weaker for PR sum and nSAT std than for nSAT mean in both periods. Also, the IMS is larger for PR sum and nSAT std than for nSAT mean. This finding suggests that there is a larger discrepancy in portraying PR sum and nSAT std in the RCM with respect to the GCM compared to nSAT mean. The correlations between CanESM2 and CRCM5 subset regions are in general significantly lower under future climate conditions ~~. The accordance of pr sum and tas std are weaker than for tas meanin both time frames~~compared to the historical ones, apart from nSAT mean in BY, PR sum in SE and nSAT std in BY (see text in Fig. 11). For nSAT std a shift of the distribution of r towards slightly larger values is visible. All variables exhibit a future IMS increase, though not all subset regions are affected (see e.g. nSAT mean BY or nSAT std SE in Fig. 11). This suggests that under future climate conditions a considerable reduction of GCM–RCM co-variability needs to be taken into account, at least for PR sum and (weaker) for nSAT mean.

4 Discussion

4.1 General performance of the model chain

The ClimEx climate data ensemble is able to reproduce an NAO-like pattern with realistic temporal and spatial characteristics over the North Atlantic and corresponding response patterns in CEUR. ~~Coincidentally, the index derived from one realization shows a very strong temporal correlation with the REF NAO index (see Fig. A2, first and second row)~~ Ensemble mean information aggregates several realizations and thus differences towards the single REF realization may occur. However, results showed that the REF pattern may in general be seen as being “embedded” in the RCM or GCM IMS, implying that GCM, RCM and REF share comparable climate statistics.

Regarding temperature, Europe is commonly seen as divided into a region with positive NAO–response correlations in the north and negative correlations in the south (see e.g., Woollings et al., 2015) . The first is found in the here presented results, the latter is not clearly visible in the chosen domain. nSAT std is correlated negatively with the NAO, pointing towards less temperature variability in winters with positive NAO phases, and a higher variability during negative phases. Correlations of PR sums and NAO are in accordance with the prevalence of large-scale (frontal) precipitation in winter which might be affected if the large-scale circulation is altered due to NAO impulses.

~~The strong psl gradient~~ The strong SLP gradient under neutral NAO conditions over the North Atlantic noted in the CanESM2-LE though suggests an overestimation of the local atmospheric circulation with too strong westerlies~~over the North Atlantic in the background state within the CanESM2-LE~~. Similar model biases are widely reported ~~(see e.g. Ruprich-Robert and Cassou, 2015; Stephense~~ ~~Positive and negative NAO states appear to be weaker within the CanESM2-LE. This difference between GCM and REF is possibly due to the fact that REF composites were derived from 3 negative and 4 positive years whereas the GCM data provided 264 negative and 263 positive years. The GCM patterns are thus more robustly assessed (that is , less prone to incidental fluctuations of single realizations). Concerning NAO responses , they~~ (see e.g., Ruprich-Robert and Cassou, 2015; Stephenson et al., 2006;

Since the NAO index was obtained from raw SLP data, it contains the contribution of the NAO, but possibly also of micro-climatic noise or other teleconnection patterns like the East Atlantic (EA) and the Scandinavian Pattern (SCA) which interact with the NAO and exert a notable control on the North Atlantic SLP gradient (Moore et al., 2013) . Moore et al. (2013) investigated the contributions of the North Atlantic teleconnections NAO, EA and SCA in reanalysis data by separating them with empirical orthogonal functions. The authors found that the NAO accounts for about one third of winter SLP variability, and the second and third leading modes for roughly 20 % and 15 %, respectively (see also Comas-Bru and McDermott, 2014) . Thus the results presented here may be seen as representing the superposition of these atmospheric modes.

The fidelity of NAO responses further depends on two aspects: (i) the goodness of representation of the large-scale NAO-related SLP pattern in CEUR and (ii) the strength of the linear relationship between the NAO and the response variables. The first point is addressed by a good representation of the SLP pattern in RCM data (see Fig. 5). The second point may be targeted by a combination of correlation analysis of the responses and the α_1 values: NAO responses in the CEUR domain of all data sets are most reliable in regions where ~~r is significant (i.e. $|r| > 0.361$ for $p \leq 0.05$, corresponding to a coefficient of determination $r^2 = 0.130$ or 13 of tas/pr variance explained by the NAO). On the other hand, lower correlation values ($|r| < 0.361$)~~ a strong linear relationship between the NAO and the response variable may be assumed. This may be the case if the correlation coefficient between the NAO index and the variable time series on the given grid cells is significantly different from zero. Linearity though does not apply under all conditions. For example particularly strong negative NAO phases with low-ice conditions in the Arctic coincide with cooling in Europe that is weaker than expected from a linear relationship due to an accompanying warming over Siberia (Screen, 2017) . Low correlation values may also suggest that climate variability ~~at the local scale evolves differently from the global teleconnection.~~ in these regions is only to a small fraction influenced by the NAO. In these cases, the NAO as expressed by the North Atlantic SLP gradient in this study is not the most important contributor and ε_Y in Eq. (~~2~~4) is dominant. ~~Since the index was obtained from raw psl data, it contains the NAO contribution, but possibly also of other teleconnection patterns and noise.~~

Historical α_1 values (all data sources) are generally in accordance with observed composite anomalies (see also Fig. ~~??~~A4), but most so in regions with significant r. Thus, the future change of ~~tas and pr~~ nSAT and PR per unit index change is most valid where r is ~~significant~~ high and where the signals of α_1 ~~and r emerge from the internal noise~~emerge from internal variability, i.e. ~~the SNR is larger than 1~~SNR > 1. Of course, α_1 and composite maps are not identical, as on the one hand the ~~mean~~ average index value that accompanies ~~tas and pr~~ nSAT and PR anomalies is not the same (± 1 for α_1, but $+1.498$ and -2.103 for REF composites, see Fig. ~~7~~9). On the other hand, α_1 estimates a change which is singularly generated by the NAO index in a linear relationship, while composite maps originate from raw data which might include further influences (Eq. (~~2~~4)).

~~Regarding temperature, Europe is commonly seen as divided into a region with positive NAO–response correlations in the north and negative correlations in the south (see e.g., Woollings et al., 2015) . The first is found in the here presented results, the latter is not as clear in the chosen domain. Strong correlations of pr sums and NAO are in accordance with the prevalence of large-scale (frontal) precipitation in winter, which might be affected if the large-scale circulation is altered due to NAO impulses. The third variable, tas std, is correlated negatively with the NAO, pointing towards less temperature variability in winters with positive NAO phases, and a higher variability in negative phases. This also fits the observation of altered jet stream and storm track behaviour. NAO~~

4.2 Nesting approach

NAO response patterns are similar within the CanESM2-LE and CRCM5-LE, but some deviations remain due to differences in model parameterization and spatial resolution. Another possible explanation could be that the control exerted by CanESM2 through the CRCM5 ~~lateral boundary conditions (LBC)~~ LBC is insufficient, but this is unlikely given the relatively small CRCM5 domain implying stronger LBC control (Leduc and Laprise, 2009), in addition to the strong spectral nudging of large scales that was applied in the production of the CRCM5-LE (Leduc et al., 2019). Also, the large-scale SLP pattern over CEUR shows no large errors in the CRCM5-LE with respect to its driving data sources (see Fig. 5). Nevertheless, the influence of the lateral boundary conditions appears to vary over the CRCM5 domain, being a bit weaker in the southern part. It is worth noting that this feature is less pronounced when CRCM5 is driven by ERA-I as compared with CanESM2, highlighting the importance to investigate further the interactions between global atmospheric circulation, surface forcings (e.g., topography and land-sea contrasts) and local feedbacks. The CRCM5 reproduces the ~~structure found in ERA-I~~ REF response structures much finer than the CanESM2 and adds some robust high resolution geographical features which are clearly visible within the ensemble mean.

~~Typical~~ Apart from the coarser pattern resolution, there is also a shift in the spatial climate patterns in the CanESM2-LE within the CEUR domain with respect to ERA-I data which is not found in the CRCM5-LE: for example, typical continental climate features, such as ~~higher tas~~ high nSAT variability (as indicated by Fig. ~~4~~6), are shifted southwards in the CanESM2-LE with respect to CRCM5-LE data (or ERA-I). This shift may be explained by the fact that due to coarser spatial resolution the GCM topography shows land grid cells where the Mediterranean ~~extends~~ or the Baltic Sea extend in ERA-I and CRCM5; thus, in the GCM, the continent Europe also occupies a region ~~,~~ which is sea in ERA-I. Assuming that the land–sea distribution affects the climate evolution, the GCM also experiences a geographical shift of climatic characteristics (such as continental properties) compared with the ERA-I ~~data. This in turn suggests~~ and RCM data within the study domain. Another example is the dividing line for NAO–PR sum relations (see Fig. 8) which shows a displacement in the GCM compared to the RCM. This displacement is related to the GCM orography which deviates due to the coarser spatial resolution in shape, position and height from the RCM orography. These findings suggest that similar responses of GCM and RCM to the NAO may not be visible at the same geographical location (i.e. ~~,~~ coordinates), but under similar geographical conditions (exposition, altitude, distance to sea). Continuing this thought, the RCM reproducing the spatial climatic patterns in the “correct” location is another expression of the RCM added value for regional or local scale analyses. However, for general statements on this issue, analyses on a larger domain would be necessary.

~~As mentioned previously~~On the regional scale, the correlations in the CRCM5 are significantly stronger in several regions than in the CanESM2 ~~.~~ (see Fig. 6–8). These are not evened out by spatial aggregation (see Fig. ~~8)and the weaker correlations in CanESM2-LE mean cannot be related to a larger inter-member spread in the CanESM2-LE in general (see Figs. 4, 5 and 6 (d), (f)).~~ 10). Thus, in the CRCM5-LE, more variance is explained by the NAO (i.e. ~~,~~ by large-scale circulation) than in the CanESM2-LE. ~~Figure A3 shows that the CRCM5 tends to underestimate (overestimate) average winter mean tas in the northern (southern) part of the domain, regardless of the driving data (see first column for ERA-I and third column for CanESM2), whereas winter pr sums are overestimated nearly in the entire domain with strongest values in the south-eastern part. The GCM is also found to overestimate temperature and precipitation in the entire domain.On the other hand, explained~~ Explained variance is higher in the single realizations of ERA-I and CRCM5/ERA-I than in the ensemble mean of GCM and RCM. ~~Interestingly, Stephenson et al. (2006) , who worked on a multi-model ensemble (with each model represented by a~~

~~single member), noted a tendency for overestimation of explained variance regarding temperature and precipitation compared to observations in all of their models~~

4.3 Internal Variability

In general, the 50 NAO signals from the atmospheric “inflow” as given by the GCM boundary conditions are correctly translated into 50 regional responses of the RCM regarding the range of internal variability.

The large ensemble internal variability favours a smoothing of structures in the ensemble ~~average~~mean. However, as the ensemble mean (GCM and RCM) reproduces patterns very similar to the observed ones, the atmospheric dynamics behind ~~are correctly incorporated~~ can be regarded as correctly reproduced in all members~~: each pattern, as given by a single member, may occur as a response to the NAO, even if it deviates strongly from the observed patterns~~.

~~In general, the 50 NAO signals from the atmospheric “inflow” as given by the GCM boundary conditions are correctly translated into 50 regional responses of the RCM regarding the range of internal variability . The~~ When looking at spatially explicit ensemble sd maps (see Figs. 6–7 and A3), the RCM LE exhibits higher ensemble sd values than the GCM. This is in accordance with Giorgi et al. (2009) who stated that internal variability at finer scales tends to be larger compared to larger scales. However, the amplitude of the ~~inter-member spread~~ IMS of r ~~is similar in~~ in the aggregated RCM and GCM subset regions is similar. Thus, the range of internal variability regarding the strength of the NAO–response relationship is transferred during nesting and the CRCM5 added internal variability (Leduc et al., 2019) does not significantly alter it. However, the ~~spread is shifted towards~~ ensemble values are shifted towards significantly higher r values in the RCM compared to the GCM ~~. Similar results are found when~~ in both time frames, but not in the SE region.

When comparing present and future values~~:~~ , a vertical shift of the boxes in Fig. ~~8~~ 10 indicates that r is reduced in the future, but the inter-quartile distance of the r distributions (box size) stays nearly the same ~~.~~for GCM and RCM. This shows that the uncertainty range of the signals does not change significantly in the future horizon.

~~The REF time series fits the ensemble statistics and may therefore be seen as a realization out of the ensemble.In other words~~Temporally constant or only negligibly varying internal variability was already found for global mean temperature in Hawkins and Sutton (2009) and assumed for global mean precipitation in Hawkins and Sutton (2011) . With the here presented results, it can also be argued that ~~the ensemble and the REF time series represent the same climate statistics.The ensemble thus produces NAO and response patterns that are more robust than patterns of single realizations.~~ internal variability of more complex parameters (such as the NAO–response relationship quantified via Pearson correlation) shows no siginificant changes between historical and future periods.

~~Further on, the std.dev50 maps in Figs.4, 5 and 6 show that the deviation of tas mean, tas std and pr sum~~ When looking at the spatial distribution of α_1 ~~also have a spatial dimension: generally, this deviation is strongest in regions with highest change in CRCM5-LE, but not automatically in CanESM2-LE~~sd however, several regions show slight future increases or decreases which are not necessarily consistent between GCM and RCM. Also, the potential time-dependent evolution of the IMS in the course of the analyzed periods is not taken into account.

It has to be added that this study evaluated two 30-year blocks rather than continuous time series, treating the NAO–response relationship as stationary during these blocks such that the IMS of both periods represents generalized conditions for 1981–2010 and 2070–2099. According to Comas-Bru and McDermott (2014) , potential non-stationarity in NAO–response relationships can at least partly be attributed to influences of the EA/SCA patterns on the NAO, and especially the geographical position of the North Atlantic SLP gradient.

~~The NAO index variability is~~

4.4 Climate Change

The results showed that historical and projected future climate statistics deviate such that the comparison of relationships in both periods remains difficult: the NAO pattern changes, NAO index variability and nSAT and PR responses are reduced in the future climate simulation~~at the same time as tas variability. This reduction is accompanied by a higher inter-member spread of α_{T}~~.

The relative prevalence of negative index phases occurs in correspondence to a generally strengthened high pressure ridge over the North Atlantic and especially Greenland (see Fig. ~~3~~ 4 (g)). The latter feature is supposed to ~~foster~~ be related with the emergence of negative index phases ~~(Hanna et al., 2015; Benedict et al., 2004) .~~ (Hanna et al., 2015; Woollings et al., 2010; Gillett and Fyf Another relationship ties the emergence of negative NAO index phases to reduced sea ice extents: Warner (2018) found that particularly October sea ice extent over the Barents/Kara Sea is positively correlated with the NAO in that it leads to strengthened IL and AH. Consequently, a reduced sea ice extent leads to negative NAO phases, but this relationship is not simply linear (Warner, 2018) . For example, Screen (2017) note that negative NAO events tend to be stronger during winters with low sea ice extents. The NAO-sea ice relationship may follow from sea ice effects on the stratospheric polar vortex or from tropospheric Arctic amplification which reduces the meridional temperature gradient leading to a weakened, more wavy jetstream in the mid-latitudes (Warner, 2018) . The CanESM2-LE is known to show a low bias regarding Arctic sea ice in all seasons compared to observations (Kushner et al., 2018) , but it follows quite correctly the observed downward trend (Kirchmeier-Young et al., 2017) and leads to a clear reduction of sea ice in the 2070–2099 horizon compared to 1981–2010 in the entire Arctic and also the Barents/Kara Sea as was verified with the CanESM2 variable “sea ice concentration” (not shown).

An increasing frequency (relative to positive phases) of negative NAO events as noted in Fig. ~~2~~ 9 favours more cold and harsh winters in theory due to the advection of continental Eurasian air masses (Screen, 2017) which is in great contradiction to ~~the~~ projected future background conditions (warmer, moister, see Leduc et al., 2019) that would rather, likewise following from theory, accompany positive phases. On the other hand, the response to NAO impulses is clearly reduced ~~in tas mean, pr sum and tas stdα_{T}~~for nSAT mean, PR sum and nSAT std. A coherent explication for this discrepancy might be that as correlations weaken, the Eurasian ~~(continental) influence~~ influence (advection of cold, dry airmasses) during negative phases may be repressed or weaker in its occurrence than now or, as indicated by Screen (2017) , is actually increasing warmer air mass advection. As less ~~tas and pr~~ nSAT and PR variance is explained by the NAO in the future climate projections than in the historical period, the influence of this climate mode on CEUR climate ~~is~~ may be seen as potentially reduced.

5 Conclusions

In this study, a RCM single-model initial condition large ensemble was analyzed with a special focus on the downscaled responses to a teleconnection, the NAO, that is present in the driving data. For proper assessment, the driving GCM ensemble was also included in the study. Referring to the key questions raised in the introduction, it can be stated that:

(a) Both large ensembles within the ClimEx project climate model chain are able to depict a robust ~~, realistic~~ NAO pattern under current forcing conditions~~, although the NAO structure itself evolves outside the analyzed CEUR domain (key question (a))~~. Each member represents a distinct climate evolution while sharing comparable statistics with all other 49 realizations ~~. The~~ and producing NAO and response patterns that are more robust than patterns of single realizations. The ensemble also shows comparable climate statistics with the REF time series and patterns. The clearly visible connection of the NAO with ~~tas mean and pr~~ nSAT mean and PR sum follows well-known patterns~~, but also the influence on tas~~. The influence of the NAO on nSAT variability, as expressed by the analyses on ~~tas~~ nSAT std, is ~~remarkable.~~ also remarkable.

(b) The RCM is able to reproduce the large-scale SLP pattern and realistic response patterns in the analyzed domain. Clearly more topographic features are visible in the CRCM5-LE than in the CanESM2-LE which suggests added value of the RCM regarding the evaluation of small-scale NAO impacts. Deviations of nSAT and PR responses between members vary spatially within the domain and are found mostly in regions with strongest NAO responses.

(c) Internal variability of the NAO pattern is expressed very well within the 50 member single-model ensemble, and easily spans the observations regarding various indicators~~, like the strength of the teleconnection or the temporal variability of the index time series in a 30 year period.~~ . The range of NAO responses is ~~transferred correctly from~~ represented consistently between the driving GCM ~~into~~ and the nested RCM. The spread is shifted towards stronger ~~NAO–tas~~NAO–nSAT/~~pr~~ PR relations in the RCM compared to the GCM in both time horizons. ~~Clearly more topographic features are visible in the CRCM5-LE than in the CanESM2-LE which proves the added value of RCM regarding the evaluation of regional NAO impacts. Deviations of NAO tas and pr responses between members vary spatially within the domain and are found mostly in regions with strongest NAO responses (key questions (b), (c)).~~

(d) Concerning climate change~~(key question (d))~~, several changes go hand in hand: the winter index variability is reduced, the overall winter variability of ~~tas and pr~~ nSAT and PR and also the fraction of NAO-explained ~~tas~~ nSAT is reduced, the relationship between NAO and ~~climate is weakenedand~~ response variables is weakened, the RMS* error regarding the large-scale SLP pattern between GCM and RCM slightly increases, and the co-variability of CanESM2 and CRCM5 subset regions for all weather variables is reduced.

While these results are especially valid for the ClimEx data sets, they allow drawing some general conclusions. The results strengthen the validity of the climate module for further applications, as important large-scale teleconnections only present in the GCM propagate properly to the ~~fine scale~~ fine-scale dynamics in the RCM. ~~Thus the~~ The RCM does not alter the spread of driving GCM data which is a valuable information for impact modelling with a focus on internal variability. The results also stress the importance of single-model ensembles for evaluating and estimating internal variability since single realizations show considerable variations among themselves and also deviations from the ensemble mean. So the ensemble mean and the ensemble spread together are needed for robust assessment of climate modes and whether a given model is able to reproduce the phenomenon of interest.

Data availability. Data used in this study may be retrieved from the following sources:

CanESM2-LE data is available via https://open.canada.ca/data/en/dataset/aa7b6823-fd1e-49ff-a6fb-68076a4a477c.

CRCM5-LE data can be retrieved at https://climex-data.srv.lrz.de/Public/

ERA-Interim Reanalysis data set was obtained at https://apps.ecmwf.int/datasets/data/interim-full-daily/levtype=sfc/

Appendix A

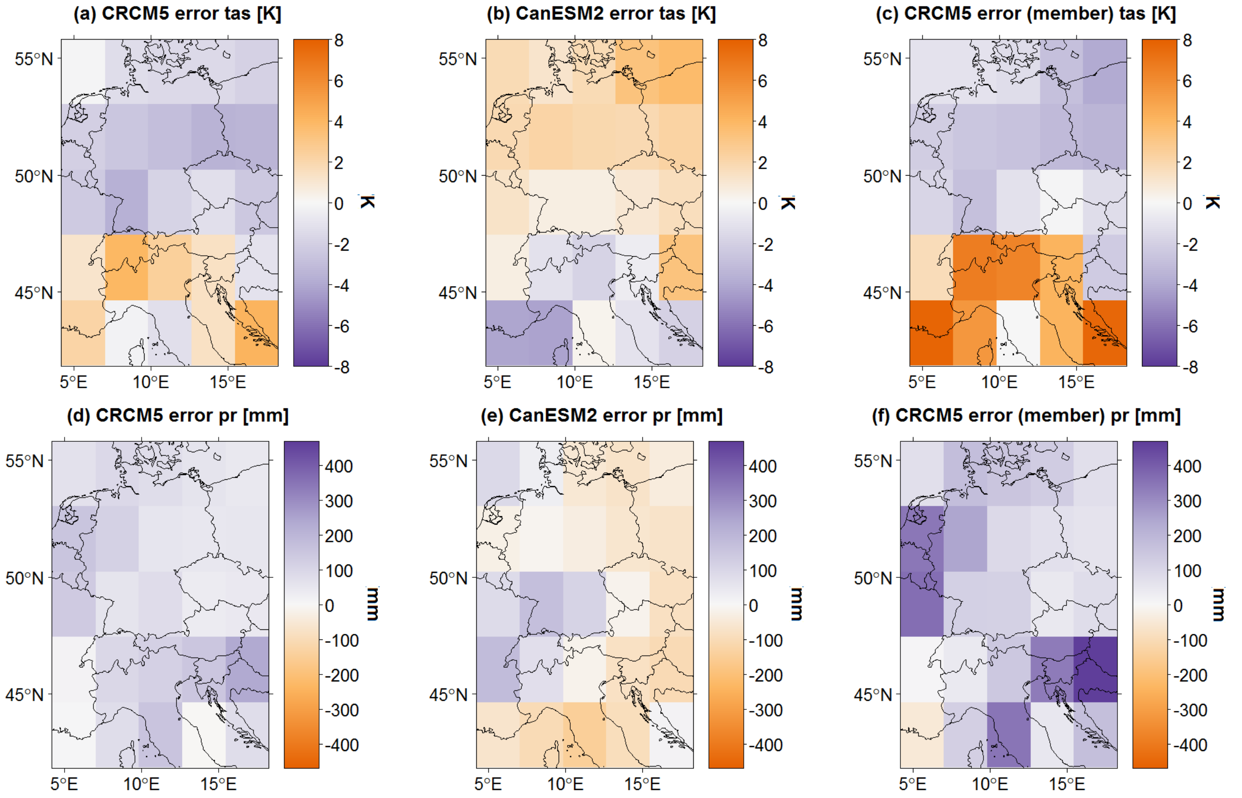

Figure A1. ~~ERA-I anomalies from the long-term~~ Model errors for 30 winter mean ~~of tas~~ nSAT mean ~~in~~ ((a)–(c)) and ~~pr~~ 30 winter mean PR sum ((d)–(f)) in ~~in NAO positive~~ GCM resolution (~~1989, 1990, 1994, 1995~~2.8°)~~and negative~~ . First column: error of CRCM5 under "perfect"/ERA-I boundary conditions (~~1996, 2001, 2010~~difference between CRCM5/ERA-I and ERA-I)~~winters~~. ~~Mean index value for positive~~ Second column: error of GCM towards ERA-I data (~~negative~~ensemble mean of differences between GCM members and ERA-I)~~NAO phases is +1.498~~ . Third column: CRCM5 error under GCM boundary conditions (~~−2.103~~ensemble mean of differences between CRCM5 members and corresponding CanESM2 members).

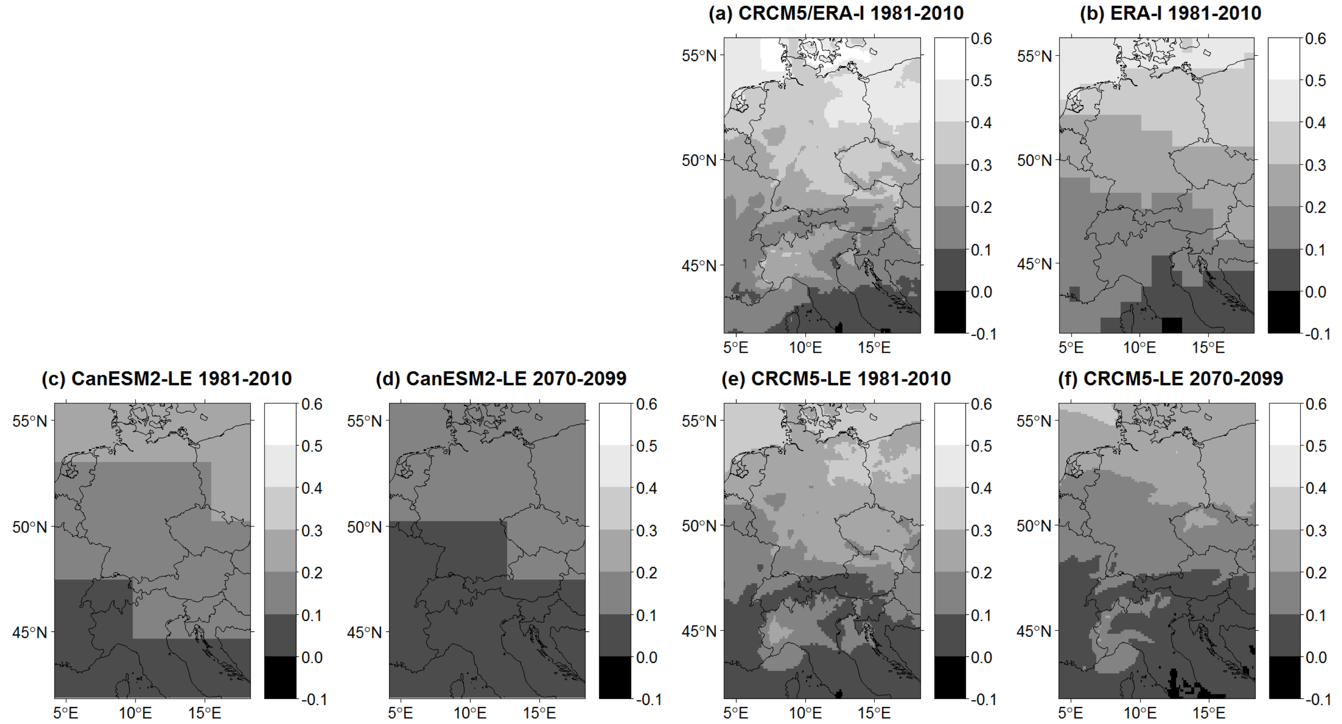

Figure A2. Ratio of ~~tas~~ nSAT α_1 and winter mean daily standard deviation of ~~tas~~nSAT for CRCM5/ERA-I (a) and ERA-I (b) under historical conditions and CanESM2-LE mean ((c)–(d)) and CRMC5-LE mean ((e)–(f)) under historical and future climate conditions. The panels show the fraction of nSAT α_1 on winter mean daily standard deviation of nSAT.

~~Signal to noise ratio (SNR) between ensemble mean and std.dev50 for tas mean: Pearson correlation coefficient ((a)–(d)) and α_1 value ((e)–(h)). Black: regions with SNR $<|1|$. Negative values emerging due to negative r and α_1 values.~~

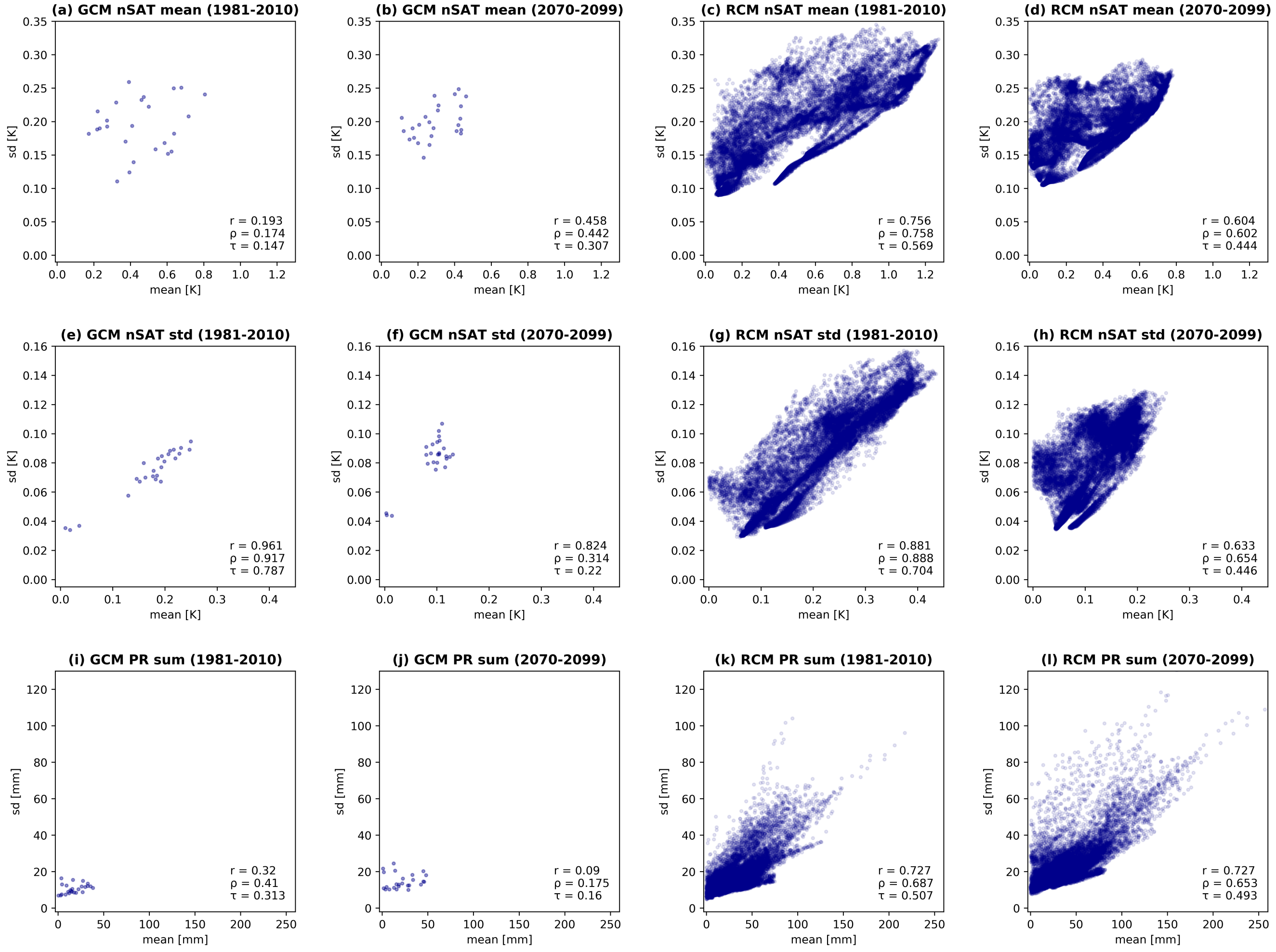

Figure A3. ~~Signal to noise ratio (SNR) ~~Relationship between ~~ensemble ~~LE mean and ~~std.dev50 ~~sd values for ~~pr sum: Pearson correlation coefficient ~~variables nSAT mean (~~(~~a)–(d)~~) and α_1 value ~~, nSAT std (~~(~~e)–(h), PR sum (i)–(l) for hist and fut time frames. ~~Black~~Upper right corner: ~~regions with SNR $<|1|$. Negative values emerging due to negative ~~r~~ and α_1 values~~ – Pearson correlation coefficient, ρ – Spearman rank correlation coefficient, τ – Kendall's Tau.~~Usage of symmetric colour bar as only strength of SNR is of importance in this case. Model errors for 30 winter tas mean averages ((a)–(c)) and 30 winter pr sum averages ((d)–(f)) in GCM resolution (2.8). First column: error of CRCM5 under "perfect"/ERA-I boundary conditions (difference between CRCM5/ERA-I and ERA-I). Second column: anomaly of GCM component of RCM towards ERA-I data (ensemble average of differences between GCM component of RCM and ERA-I). Third column: CRCM5 error under GCM boundary conditions (ensemble average of differences between CRCM5 members and corresponding CanESM2 members).~~

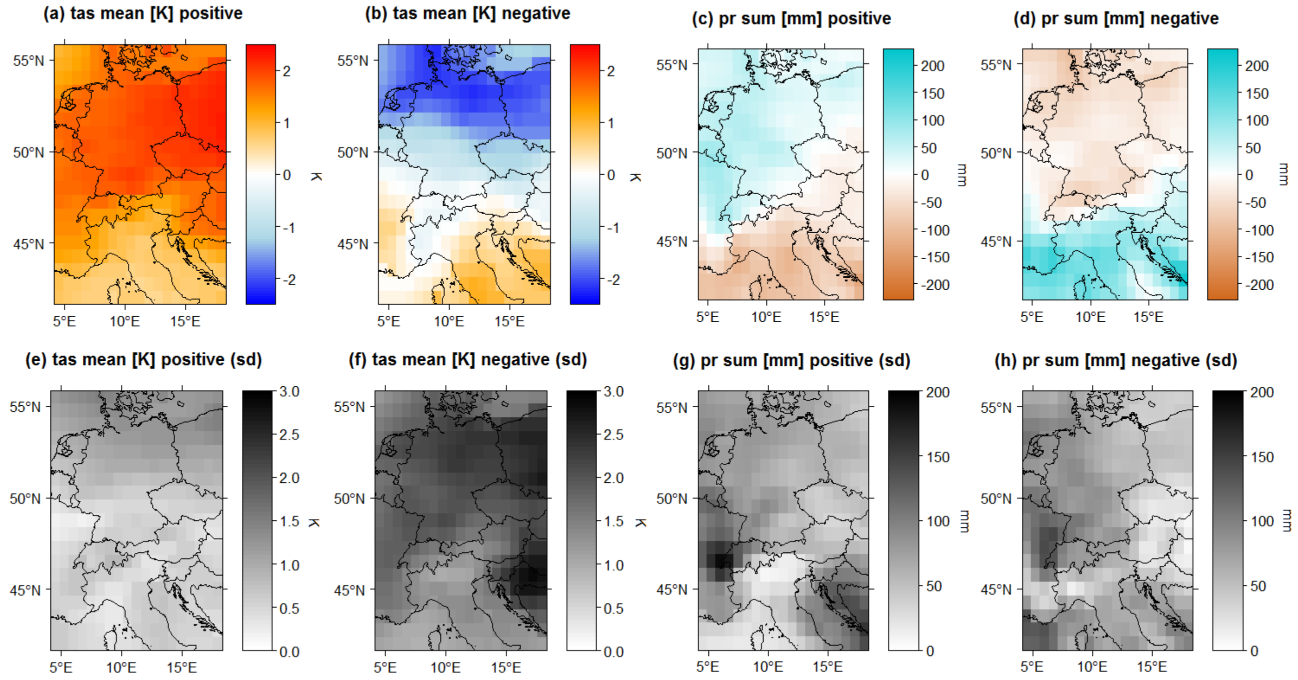

Figure A4. ERA-I anomalies from the long-term mean of nSAT mean in [K] and PR sum in [mm] in NAO positive (1989, 1990, 1994, 1995) and negative (1996, 2001, 2010) winters. Mean index ~~time series~~ value for ~~all 50 CanESM2 members~~ positive (~~see row names on the left side~~negative) ~~for 1981–2010. The first row presents the reference~~ NAO phases is +1.498 (~~ERA-I~~−2.103)~~time series~~.~~Colour: winter index values of single years.~~

Author contributions. This study was conceptualized by AB under supervision of RL. Formal analysis, visualization of results and writing of the original draft was performed by AB. All authors contributed to the interpretation of the findings and revision of the paper.

Competing interests. The authors declare that they have no competing interests.

Acknowledgements. The authors would like to thank two anonymous reviewers for their valuable input on this work. The production of ClimEx was funded within the ClimEx project by the Bavarian State Ministry for the Environment and Consumer Protection. The CRCM5 was developed by the ESCER centre of Université du Québec à Montréal (UQAM; www.escer.uqam.ca) in collaboration with Environment and Climate Change Canada. We acknowledge Environment and Climate Change Canada's Canadian Centre for Climate Modelling and Analysis for executing and making available the CanESM2 Large Ensemble simulations used in this study, and the Canadian Sea Ice and Snow Evolution Network for proposing the simulations. Computations with the CRCM5 for the ClimEx project were made on the SuperMUC supercomputer at Leibniz Supercomputing Centre (LRZ) of the Bavarian Academy of Sciences and Humanities. The operation of this supercomputer is funded via the Gauss Centre for Supercomputing (GCS) by the German Federal Ministry of Education and Research and the Bavarian State Ministry of Education, Science and the Arts.

References

Andrews, M., Knight, J., and Gray, L.: A simulated lagged response of the North Atlantic Oscillation to the solar cycle over the period 1960–2009, Environmental Research Letters, 10, 054 022, https://doi.org/10.1088/1748-9326/10/5/054022, https://doi.org/10.1088%2F1748-9326%2F10%2F5%2F054022, 2015.

Benedict, J., Lee, S., and Feldstein, S.: Synoptic View of the North Atlantic Oscillation, Journal of the Atmospheric Sciences, 61, 121–144, https://doi.org/10.1175/1520-0469(2004)061<0121:SVOTNA>2.0.CO;2, https://doi.org/10.1175/1520-0469(2004)061<0121:SVOTNA>2.0.CO;2, 2004.

Cattiaux, J., Douville, H., and Peings, Y.: European temperatures in CMIP5: origins of present-day biases and future uncertainties, Climate Dynamics, 41, 2889–2907, https://doi.org/10.1007/s00382-013-1731-y, https://doi.org/10.1007/s00382-013-1731-y, 2013.

Christensen, J., Krishna Kumar, K., Aldrian, E., An, S.-I., Cavalcanti, I., de Castro, M., Dong, W., Goswami, P., Hall, A., Kanyanga, J., Kitoh, A., Kossin, J., Lau, N.-C., Renwick, J., Stephenson, D., Xie, S.-P., and Zhou, T.: Climate Phenomena and their Relevance for Future Regional Climate Change, in: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, edited by Stocker, T., Qin, D., Plattner, G.-K., Tignor, M., Allen, S., Boschung, J., Nauels, A., Xia, Y., Bex, V., and Midgley, P., pp. 1271–1308, Cambridge University Press, Cambridge, United Kingdom and New York, USA, 2013.

Comas-Bru, L. and McDermott, F.: 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, https://doi.org/10.1002/qj.2158, https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.2158, 2014.

Cropper, T., Hanna, E., Valente, M., and Jónsson, T.: A daily Azores-Iceland North Atlantic Oscillation index back to 1850, Geoscience Data Journal, 2, 12–24, https://doi.org/10.1002/gdj3.23, 2015.

Dee, D., Uppala, S., Simmons, A., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hersbach, H., Hólm, E., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A., Monge-Sanz, B., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., and Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, Quarterly Journal of the Royal Meteorological Society, 137, 553–597, https://doi.org/10.1002/qj.828, 2011.

Delworth, T., Zeng, F., Vecchi, G., Yang, X., Zhang, L., and Zhang, R.: The North Atlantic Oscillation as a driver of rapid climate change in the Northern Hemisphere, Nature Geoscience, 9, 509–512, https://doi.org/10.1038/ngeo2738, https://doi.org/10.1038/ngeo2738, 2016.

Déqué, M., Rowell, D., Lüthi, D., Giorgi, F., Christensen, J., Rockel, B., Jacob, D., Kjellström, E., de Castro, M., and van den Hurk, B.: An intercomparison of regional climate simulations for Europe: assessing uncertainties in model projections, Climatic Change, 81, 53–70, https://doi.org/10.1007/s10584-006-9228-x, 2007.

Deser, C., Phillips, A., Bourdette, V., and Teng, H.: Uncertainty in climate change projections: the role of internal variability, Climate Dynamics, 38, 527–546, https://doi.org/10.1007/s00382-010-0977-x, https://doi.org/10.1007/s00382-010-0977-x, 2012.

Deser, C., Hurrell, J., and Phillips, A.: The role of the North Atlantic Oscillation in European climate projections, Climate Dynamics, 49, 3141–3157, https://doi.org/10.1007/s00382-016-3502-z, https://doi.org/10.1007/s00382-016-3502-z, 2017.

Fyfe, J., Derksen, C., Mudryk, L., Flato, G., Santer, B., Swart, N., Molotch, N., Zhang, X., Wan, H., Arora, V., Scinocca, J., and Jiao, Y.: Large near-term projected snowpack loss over the western United States, Nature Communications, 8, https://doi.org/10.1038/ncomms14996, https://doi.org/10.1038/ncomms14996, 2017.

Gillett, N. and Fyfe, J.: Annular mode changes in the CMIP5 simulations, Geophysical Research Letters, 40, 1189–1193, https://doi.org/10.1002/grl.50249, https://doi.org/10.1002/grl.50249, 2013.

Giorgi, F., Jones, C., and Asrar, G. R.: Addressing climate information needs at the regional level: the CORDEX framework, WMO Bulletin, 58, 175–183, https://public.wmo.int/en/bulletin/addressing-climate-information-needs-regional-level-cordex-framework, 2009.

Hanna, E., Cropper, T., Jones, P., Scaife, A., and Allan, R.: Recent seasonal asymmetric changes in the NAO (a marked summer decline and increased winter variability) and associated changes in the AO and Greenland Blocking Index, International Journal of Climatology, 35, 2540–2554, https://doi.org/10.1002/joc.4157, https://doi.org/10.1002/joc.4157, 2015.

Hansen, F., Greatbatch, R., Gollan, G., Jung, T., and Weisheimer, A.: Remote control of North Atlantic Oscillation predictability via the stratosphere, Quarterly Journal of the Royal Meteorological Society, 143, 706–719, https://doi.org/10.1002/qj.2958, https://doi.org/10.1002/qj.2958, 2017.

Hawkins, E. and Sutton, R.: The Potential to Narrow Uncertainty in Regional Climate Predictions, Bulletin of the American Meteorological Society, 90, 1095–1108, https://doi.org/10.1175/2009BAMS2607.1, https://doi.org/10.1175/2009BAMS2607.1, 2009.

Hawkins, E. and Sutton, R.: The Potential to narrow uncertainty in projections of regional precipitation Change, Climate Dynamics, 37, 407–418, https://doi.org/10.1007/s00382-010-0810-6, https://doi.org/10.1007/s00382-010-0810-6, 2011.

Hurrell, J. W.: Decadal Trends in the North Atlantic Oscillation: Regional Temperatures and Precipitation, Science, 269, 676–679, https://doi.org/10.1126/science.269.5224.676, 1995.

Hurrell, J. W. and Deser, C.: North Atlantic climate variability: The role of the North Atlantic Oscillation, Journal of Marine Systems, 78, 28–41, https://doi.org/https://doi.org/10.1016/j.jmarsys.2008.11.026, 2009.

Hurrell, J. W. and Van Loon, H.: Decadal variations in climate associated with the North Atlantic Oscillation, Climatic Change, 36, 301–326, https://doi.org/10.1023/A:1005314315270, https://doi.org/10.1023/A:1005314315270, 1997.

Iles, C. and Hegerl, G.: Role of the North Atlantic Oscillation in decadal temperature trends, Environmental Research Letters, 12, 114 010, https://doi.org/10.1088/1748-9326/aa9152, https://doi.org/10.1088/1748-9326/aa9152, 2017.

Jones, P., Osborn, T., and Briffa, K.: Pressure-Based Measures of the North Atlantic Oscillation (NAO): A Comparison and an Assessment of Changes in the Strength of the NAO and its Influence on Surface Climate Parameters, in: The North Atlantic Oscillation: Climatic Significance and Environmental Impact, edited by Hurrell, J., Kushnir, Y., Ottersen, G., and Visbeck, M., pp. 51–62, American Geophysical Union (AGU), Washington, D.C., https://doi.org/10.1029/134GM03, 2013.

Kirchmeier-Young, M. C., Zwiers, F. W., and Gillett, N. P.: Attribution of Extreme Events in Arctic Sea Ice Extent, Journal of Climate, 30, 553–571, https://doi.org/10.1175/JCLI-D-16-0412.1, 2017.

Kirtman, B., Power, S., Adedoyin, J., Boer, G., Bojariu, R., Camilloni, I., Doblas-Reyes, F., Fiore, A., Kimoto, M., Meehl, G., Prather, M., Sarr, A., Schär, C., Sutton, R., van Oldenborgh, G., Vecchi, G., and Wang, H.: Near-term Climate Change: Projections and Predictability, in: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, edited by Stocker, T., Qin, D., Plattner, G.-K., Tignor, M., Allen, S., Boschung, J., Nauels, A., Xia, Y., Bex, V., and Midgley, P., pp. 953–1028, Cambridge University Press, Cambridge, United Kingdom and New York, USA, 2013.

Kushner, P. J., Mudryk, L. R., Merryfield, W., Ambadan, J. T., Berg, A., Bichet, A., Brown, R., Derksen, C., Déry, S. J., Dirkson, A., Flato, G., Fletcher, C. G., Fyfe, J. C., Gillett, N., Haas, C., Howell, S., Laliberté, F., McCusker, K., Sigmond, M., Sospedra-Alfonso, R., Tandon, N. F., Thackeray, C., Tremblay, B., and Zwiers, F. W.: Canadian snow and sea ice: assessment of snow, sea ice, and related climate processes in Canada's Earth system model and climate-prediction system, The Cryosphere, 12, 1137–1156, https://doi.org/10.5194/tc-12-1137-2018, https://www.the-cryosphere.net/12/1137/2018/, 2018.

Leduc, M. and Laprise, R.: Regional climate model sensitivity to domain size, Climate Dynamics, 32, 833–854, https://doi.org/10.1007/s00382-008-0400-z, https://doi.org/10.1007/s00382-008-0400-z, 2009.

Leduc, M., Mailhot, A., Frigon, A., Martel, J.-L., Ludwig, R., Brietzke, G., Giguère, M., Brissette, F., Turcotte, R., Braun, M., and Scinocca, J.: The ClimEx Project: A 50-Member Ensemble of Climate Change Projections at 12-km Resolution over Europe and Northeastern North America with the Canadian Regional Climate Model (CRCM5), Journal of Applied Meteorology and Climatology, 58, 663–693, https://doi.org/10.1175/JAMC-D-18-0021.1, https://doi.org/10.1175/JAMC-D-18-0021.1, 2019.

Moore, G., Renfrew, I., and Pickart, R.: Multidecadal Mobility of the North Atlantic Oscillation, Journal of Climate, 26, 2453–2466, https://doi.org/10.1175/JCLI-D-12-00023.1, https://doi.org/10.1175/JCLI-D-12-00023.1, 2013.

Osborn, T.: Simulating the winter North Atlantic Oscillation: the roles of internal variability and greenhouse gas forcing, Climate Dynamics, 22, 605–623, https://doi.org/10.1007/s00382-004-0405-1, https://doi.org/10.1007/s00382-004-0405-1, 2004.

Pokorná, L. and Huth, R.: Climate impacts of the NAO are sensitive to how the NAO is defined, Theoretical and Applied Climatology, 119, 639–652, https://doi.org/10.1007/s00704-014-1116-0, https://doi.org/10.1007/s00704-014-1116-0, 2015.

Reintges, A., Latif, M., and Park, W.: Sub-decadal North Atlantic Oscillation variability in observations and the Kiel Climate Model, Climate Dynamics, 48, 3475–3487, https://doi.org/10.1007/s00382-016-3279-0, https://doi.org/10.1007/s00382-016-3279-0, 2017.

Rogers, J.: The Association between the North Atlantic Oscillation and the Southern Oscillation in the Northern Hemisphere, Monthly Weather Review, 112, 1999–2015, https://doi.org/10.1175/1520-0493(1984)112<1999:TABTNA>2.0.CO;2, https://doi.org/10.1175/1520-0493(1984)112<1999:TABTNA>2.0.CO;2, 1984.

Ruprich-Robert, Y. and Cassou, C.: Combined influences of seasonal East Atlantic Pattern and North Atlantic Oscillation to excite Atlantic multidecadal variability in a climate model, Climate Dynamics, 44, 229–253, https://doi.org/10.1007/s00382-014-2176-7, https://doi.org/10.1007/s00382-014-2176-7, 2015.

Schulzweida, U.: CDO User Guide. Climate Data Operators Version 1.9.1. October 2017, MPI for Meteorology. URL: https://code.mpimet.mpg.de/projects/cdo/embedded/cdo.pdf., access: 01.04.2018, 2017.

Screen, J.: The missing Northern European winter cooling response to Arctic sea ice loss, Nature Communications, 8, https://doi.org/10.1038/ncomms14603, https://doi.org/10.1038/ncomms14603, 2017.

Stephenson, D., Pavan, V., Collins, M., Junge, M., and Quadrelli, R.: North Atlantic Oscillation response to transient greenhouse gas forcing and the impact on European winter climate: a CMIP2 multi-model assessment, Climate Dynamics, 27, 401–420, https://doi.org/10.1007/s00382-006-0140-x, https://doi.org/10.1007/s00382-006-0140-x, 2006.

Ulbrich, U. and Christoph, M.: A shift of the NAO and increasing storm track activity over Europe due to anthropogenic greenhouse gas forcing, Climate Dynamics, 15, 551–559, https://doi.org/10.1007/s003820050299, https://doi.org/10.1007/s003820050299, 1999.

Ulbrich, U., Pinto, J. G., Kupfer, H., Leckebusch, G. C., Spangehl, T., and Reyers, M.: Changing Northern Hemisphere Storm Tracks in an Ensemble of IPCC Climate Change Simulations, Journal of Climate, 21, 1669–1679, https://doi.org/10.1175/2007JCLI1992.1, http://dx.doi.org/10.1175/2007jcli1992.1, 2008.

von Storch, H. and Zwiers, F.: Statistical Analysis in Climate Research, Cambridge University Press, Cambridge, 2003.

von Trentini, F., Leduc, M., and Ludwig, R.: Assessing natural variability in RCM signals: comparison of a multi model EURO-CORDEX ensemble with a 50-member single model large ensemble, Climate Dynamics, 53, 1963–1979, https://doi.org/10.1007/s00382-019-04755-8, https://doi.org/10.1007/s00382-019-04755-8, 2019.

Warner, J. L.: Arctic sea ice – a driver of the winter NAO?, Weather, 73, 307–310, https://doi.org/10.1002/wea.3399, https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/wea.3399, 2018.

Wilks, D. S.: “The Stippling Shows Statistically Significant Grid Points”: How Research Results are Routinely Overstated and Overinterpreted, and What to Do about It, Bulletin of the American Meteorological Society, 97, 2263–2273, https://doi.org/10.1175/BAMS-D-15-00267.1, https://doi.org/10.1175/BAMS-D-15-00267.1, 2016.

Woollings, T., Hannachi, A., Hoskins, B., and Turner, A.: A Regime View of the North Atlantic Oscillation and Its Response to Anthropogenic Forcing, Journal of Climate, 23, 1291–1307, https://doi.org/10.1175/2009JCLI3087.1, https://doi.org/10.1175/2009JCLI3087.1, 2010.

Woollings, T., Franzke, C., Hodson, D., Dong, B., Barnes, E., Raible, C., and Pinto, J.: Contrasting interannual and multidecadal NAO variability, Climate Dynamics, 45, 539–556, https://doi.org/10.1007/s00382-014-2237-y, https://doi.org/10.1007/s00382-014-2237-y, 2015.

Xu, T., Shi, Z., Wang, H., and An, Z.: Nonstationary impact of the winter North Atlantic Oscillation and the response of mid-latitude Eurasian climate, Theoretical and Applied Climatology, 124, https://doi.org/10.1007/s00704-015-1396-z, 2015.

Zwiers, F. and von Storch, H.: On the role of statistics in climate research, International Journal of Climatology, 24, 665–680, https://doi.org/10.1002/joc.1027, 2004.