Responses to Reviewer Comments (Major Revision, 2^{nd} round)

Dear reviewers, dear editor,

We would like to thank you for a second round of constructive advice and suggestions. During the revision process we tried to incorporate these where possible. 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) Revised manuscript with track changes (with respect to previous version)

We answer the general remarks stated in the introducing paragraphs of each report before the point-by-point responses.

Changes in the revised manuscript that do not apply to specific comments include:

- Fig. 4: We removed the boxes indicating the AH/IL and CEUR regions to improve readability. The information on their locations is provided in Fig. 2.
- Figs. 6—8: Since the grey lines, representing the correlations between nSAT/PR and the NAO index, are difficult to read in most regions, we changed their colors to red shadings. This also helps to separate the lines from the stippling.
- Figs. A2—A4: The subpanels in Fig. A2 were re-arranged and the subpanel headers of Figs. A3 and A4 were corrected (variable names).
- Table 1: since psl of both RCM data sets was used for the calculation of RMSD, this variable is now included in this table (for some reason, the track changes did not capture it. It is included in the revised manuscript).

In the following, comments raised by the referees are marked in blue, responses in black (explanations in italics). Line numbers refer to the revised manuscript.

With kind regards,

Andrea Böhnisch on behalf of all co-authors

11 May 2020

Reviewer No. 1

This is my second review of the manuscript "Using a nested single-model large ensemble to assess the internal variability of the North Atlantic Oscillation and its climatic implications for Europe" by A. Böhnisch et al. I appreciate the amount of work the authors put into improving the manuscript. The authors adequately addressed most reviewer comments (the exception will be discussed below). In my opinion, they have indeed produced a much improved version of the manuscript: it is for the most part very clear, easy to follow, and continues to present an interesting piece of research. I enjoyed this review as I learned a lot. I would like to thank the authors for that.

I suggest publishing the paper in Earth System Dynamics after some minor revisions that I outline below have been addressed.

While the text and reasoning are much easier to follow in the current version of the manuscript than they were in the previous version, some (easily amendable) issues remain. Firstly, the authors still do not consistently use present or past tense (my previous review, comment on l. 140). While I can see they made an effort for a consistent use of past tense, this is not applied everywhere (e.g. l. 105 "Model data is compared..."; l. 176 "The NAO is quantified..."). In fact, most of the results section is written in present tense. I suggest using one or the other, and doing so consistently. My personal preference is usage of present tense for the study at hand and past tense for all cited literature. This will strongly improve readability of the paper. Second, the authors use different abbreviations for "standard deviation": sometimes it is "std" (e.g. l. 129), other times "sd" is used (e.g. l. 354). I suggest choosing one ("std"?) and sticking to it.

Use of tenses:

When re-reading the manuscript, we decided to switch to present tense in section 2 (data & methods) whenever we refer to our present study, while past tense now marks results from cited literature. In doing so, we intend to keep coherence with the results section. This follows the reviewer's suggestion and will hopefully avoid any misunderstanding regarding the presentation of our results and findings.

Abbreviations for standard deviations:

Originally, both abbreviations referred to different aspects: "std" as in "nSAT std" is a standard deviation along the time axis of temperature. "sd" in turn is the across-member standard deviation. However, for the sake of clearness we may refrain from one of the abbreviations and stick to the other one (i.e., sd). We therefore changed nSAT std to nSAT sd.

Detailed comments

ll. 13-14 How do the results stress the importance of SMILEs for estimating internal variability?

We included an explaining phrase in this sentence and changed the word "importance" to "value":

II. 13-14: „The results stress the value of single-model ensembles for the evaluation of internal variability by pointing out the large differences of NAO—response relationships among individual members."

l. 40 I think replacing "though" with ", however," would sound nicer. This is a issue several times in the document. Please consider making these changes.

Thank you. We adopted the suggestion within the revision.

l. 64 Readers that are not familiar with regional modeling would benefit from a short explanation of the term "nesting" here, I think.

We included a short parenthesis:

ll. 73-74: “Combining the driving GCM and nested RCM (i.e., driven by lateral boundary conditions of the GCM) large ensembles (LE) allows for analyzing the spread of NAO states and responses within one model chain.”

l. 67 The word "also" suggests that there were other questions asked before. As far as I can tell, there were no explicit questions up until here. I suggest omitting "also" or asking previous questions more explicitly. In light of the reference to "research questions" in line 71, I highly suggest making these research questions absolutely explicit somewhere earlier in the text.

Thank you. We now explicitly mention the research question:

ll. 77-79: “The present study targets the research 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-LE and the driven high resolution RCM-LE.“

ll. 156-166 I suggest presenting the examined 30-year long time horizons (ll. 160-166) first and then discussing their implications (ll.156-159). The current order raises several questions about the time horizons first, that are then answered in the next paragraph. By switching the order of the paragraphs, these questions could be avoided.

and

ll. 167-174 Here I have the same comment as for ll. 156-166. Please consider switching the order of the paragraphs. Otherwise the reader might wonder, e.g., how winter is defined. This is not necessary and solely an artifact of the order in which the text presents itself.

Agreed. We rearranged in both cases the paragraphs.

Equation 2 Why does the RMS have an asterisk? In the text, the "*" is confusing: I looked for a footnote several times.

The asterisk was included to indicate that this RMS is a modified RMS. But we agree that, since the metric is defined in a formula, it is not necessary and may be confusing. The asterisk was thus removed in the revised manuscript. We use the term “RMSD” now.

ll. 228-237 Recently, several references appeared in the literature using large ensemble to examine transient internal climate variability. (At least) these should be cited here: Kay et al. (2015), Sigmon & Fyfe (2016), Maher et al. (2018), Maher et al. (2019).

Thank you for these suggestions! We conducted a literature review and included references that seemed fitting (e.g. in the introduction, partly also in section 2.2.5 Addressing internal variability).

l. 249 I cannot see the "slightly low positive NAO value" surplus in figure 3a that the authors describe here. This statement does not seem essential anyway, so I suggest omitting it.

Agreed, we may omit it. This statement refers to the slight shift of curves towards positive values, i.e. crossing the horizontal 0.5-line at an index value larger than zero. The revised text says:

ll. 267-269: „The CanESM2-LE produces NAO index values which follow a distribution comparable to the ERA-I data (similar to a normal distribution with $\mu = 0$, $\sigma = 1$, Fig. 3 (a)), but the CanESM2-LE distribution appears smoother due to a larger sample size (n = 1500 for CanESM2-LE and n = 30 for ERA-I)."

All figures As it stands, stippling is really hard to see in the maps. Please redo the maps with more prominent stippling. I had trouble finding the stipples in the current version although the text told me they were there!

Thank you for this remark. We included more prominent stippling and clearly indicated in the figure captions the relevant subpanels. In order to better separate it from the isolines, we also changed their colors. See e.g. Fig. 6:

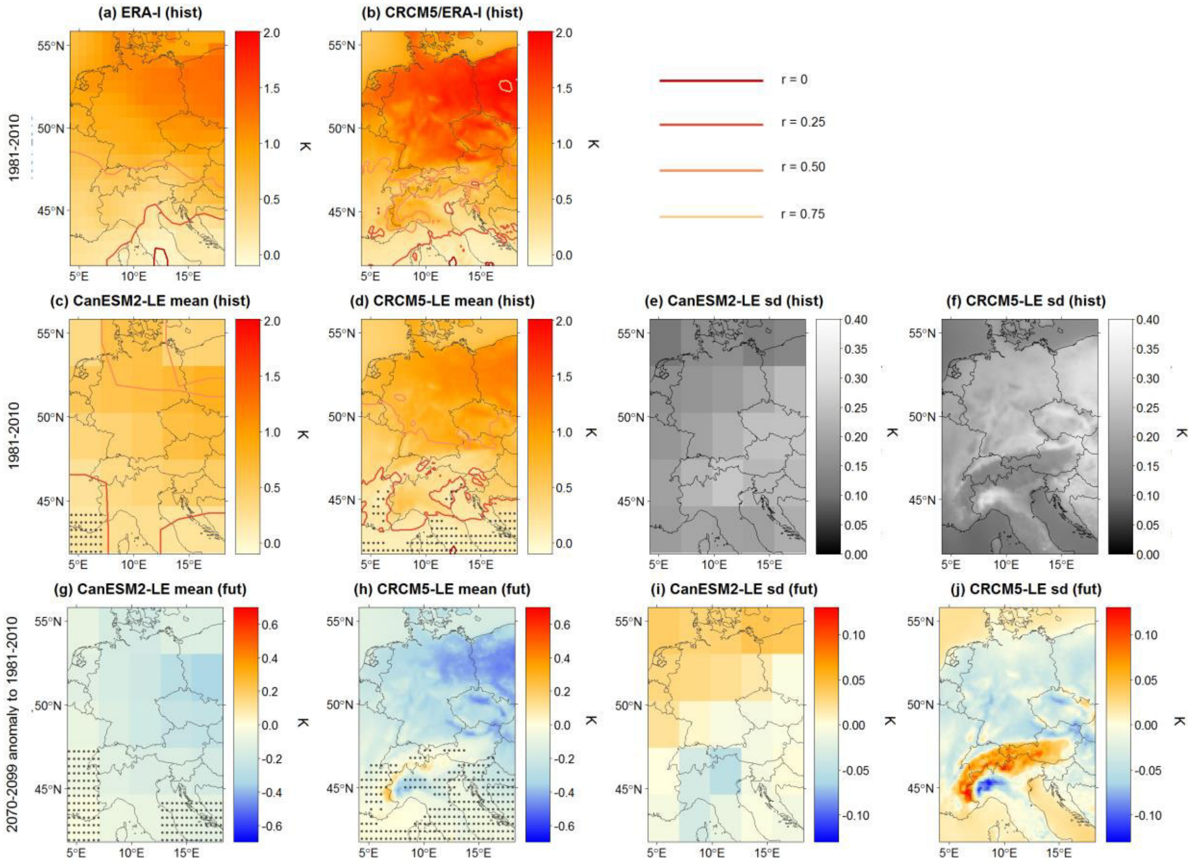


Figure 1: Spatial patterns of change in nSAT mean (α_1 in [K]) for a unit change in the NAO index for ERA-I, CRCM5/ERA-I, CanESM2-LE and CRCM5-LE in 1981–2010 ((a)–(f)) and the change in 2070–2099 with respect to 1981–2010 ((g)–(j)). Both 50-member ensembles are represented with ensemble mean ((c)–(d), (g)–(h)) and standard deviation (sd, (e)–(f), (i)–(j)) representing the inter-member spread. Reddish lines in the ensemble mean maps represent the Pearson correlation between nSAT mean and the NAO index at an increment of 0.25; red shadings see legend in upper right panel. Grey stippling in the ensemble mean maps show regions where 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. Stippling will be explained in moredetail in Section 3.2.2.

ll. 277, 282 The explanation of what stippling means comes too late here. Please first explain what stippling means, then reference it.

Thank you. We mention the stippling earlier in the text when the figure is presented.

l. 290 Please omit "is very promising as it". This statement is not necessary.

Agreed. We removed it.

Figure 5 The panel headings for (b) and (c) say "GCM", but I suspect what is shown is from the regional model. Please correct this.

This figure shows the RMSD of the large-scale SLP pattern between the regional model and its driving data: (a) between CRCM5 and ERA-I as driving data during 1981-2010 winters, (b) between CRCM5 and CanESM2 as driving data during 1981-2010 winters, and (c) like (b), but for 2070-2099 winters. As the panel headings indeed are misleading, we changed them to explicitly mention the driving data/CRCM5 combination.

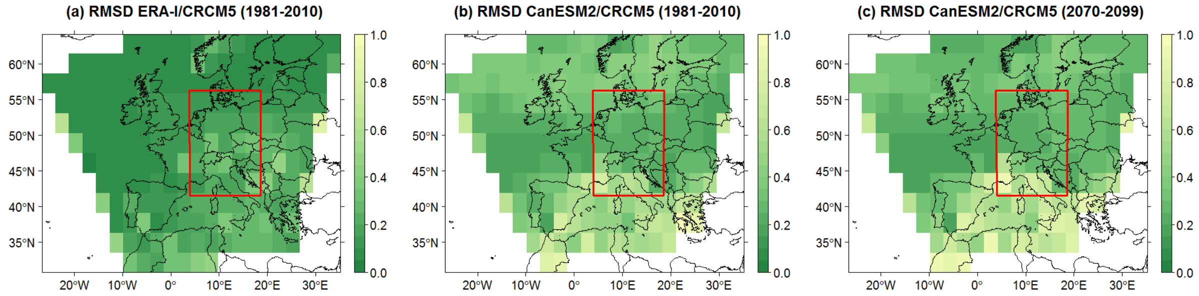


Figure 2: RMSD of monthly SLP differences between driving data and CRCM5 members, calculated following Eq. (2). Colouring: RMSD <= 1 significant at p <= 0.05 with a false detection rate smaller than 0.1 (see Wilks, 2016). (a) for driving data ERA-I (1981–2010, one realization), (b) for driving data CanESM2-LE (1981–2010, 50 members), (c) for driving data CanESM2-LE (2070–2099, 50 members). Red box: position of CEUR domain.

ll. 368ff Please use one p-value throughout the manuscript and stick to it! Using changing p-values introduces significant ambiguity into the text. For example, in line 369, what does "sometimes significant" mean? Which significance level does that refer to? Switching p-values for statistical significance gives the impression of p-hacking, and I am sure that the authors want to avoid this impression (and these results are significant and interesting enough as they are).

p-values: Thanks for your remark. In this paragraph, we wanted to show whether the variances between historical and future periods are significantly different within the future and historical ensembles. In the conclusion, we argue that they are ***not*** *since we cannot reject our null hypothesis of no differences – apart from one case with p <= 0.05 (p = 0.0443) and another one with p <= 0.1 (p = 0.0546). We agree that the deviating p-value is rather confusing and should be abandoned. So the text now only mentions the one counter-example which is significant at the level chosen for the text (i.e., p <= 0.05).*

“sometimes”: This is misleading wording. In this paragraph, it is as in the remainder of the study p <= 0.05. We rephrased to show what we actually wanted to say:

ll. 403-407: “This general finding does not change in the projected future climate: most boxes and whiskers keep their size, only GCM nSAT in the NE region is characterized by a larger range in the future (significant at p ≤ 0.05, using an F-test for comparison of variances). Some of the ensemble mean values exhibit a significant shift towards lower r values in the future for both models for nSAT mean and PR sum (see text insertions CanESM2(hist, fut) and CRCM5(hist, fut)).”

l. 374 "it is now advised" should be omitted. Who gave that advise?

Agreed, we rephrased this sentence.

l. 411: “Having analyzed GCM and RCM separately so far, the next step is to compare both ensembles”.

ll. 399-340 Replace "needs to be taken into account" with "needs to be expected"?

We rephrased, but we suggest: "needs to be considered" (l.437).

ll. 491-502 Isn't this already part 4.4?

Yes, it can be argued that this paragraph is actually part 4.4. Originally, the focus of section 4.3 was set on internal variability which is why we first put this paragraph in there. However, in order to provide a more complete image of the future conditions and changes, we include this paragraph in section 4.4 in the revised manuscript.

l. 505 This is one of the examples of surprising switch to past tense, when present tense was used before.

Adjusted. We decided to change to present tense for analyses performed in this study rather than past tense and keep it whenever results of our study are mentioned.

l. 554 What is the "climate module"? I do not think it was mentioned in the manuscript so far.

Thank you! This is a leftover of a previous revision. The climate module originally refers to one of the modules within the ClimEx project. Since we do not mention it in the manuscript any more, the term was changed to: "of this GCM-RCM combination" (l. 590).

Reference:

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, 2016.

Reviewer No. 2

In this study the regional climate model, CRCM5, is employed to dynamically downscale a 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.

In my initial review of this manuscript, I asked the authors to provide a more detailed investigation of the RCM's ability to faithfully reproduce the NAO signal in the driving data and I provided a detailed suggestion on how that might be done. The authors have done an excellent job of addressing this issue in the revised manuscript. However, in reading through the new manuscript I continue to find it difficult to read due to back-and-forth referencing (including far too many acronyms), which I mentioned in my minor comments of the first draft.

In particular, the authors have front-loaded Section 2 with discussion and derivation of diagnostics which are then used in later sections of the paper. However, the discussion in those later sections is conducted as if that earlier material was just presented. It was not, and the reader must constantly stop and hunt down the meaning of acronyms and variables. This makes it very difficult to read the paper. It could easily be resolved by bringing the motivation, derivation, and application of diagnostics and ideas together. I have identified a few places but it is problematic in more than these examples. I don't want to hold the authors to another major revision as I did not raise this so strongly in my first review. As a consequence, I offer my comments as suggestions to the authors and leave it to them to follow through. I would highly recommend that the authors attempt to improve the flow of this paper, however, as I feel it is a very good piece of work and I wouldn't want to see that obfuscated by the manner in which it was presented. With this understanding, I therefore recommend publishing this work with minor modification. My detailed comments follow.

Abbreviations/acronyms:

We went through the abbreviations in the manuscript to decide which may be abandoned. This was the case for “IMS” (inter-member spread), “LBC” (lateral boundary conditions) and “REF” (reference). On other occasions, we inserted a reminder if the abbreviation/acronym reappears several paragraphs/pages later again (this applies e.g. for the α_1 parameter, i.e. the change of temperature/precipitation associated with unit index change of the NAO). Where possible, we also tried to rephrase/rearrange phrases with excessive abbreviation use to avoid chains of acronyms.

Structure:

We understand the concerns regarding the difficulties when reading the paper due to the separation of methods and results/discussion. Nevertheless, we decided to stick to that structure and include, where applicable, reminders of the concepts in the results and discussion sections. This allows us to use the methods section for answering the question “how is this done?” for each of the key questions. However, we did remove the detailed discussion of concepts in the methods section to reduce the mentioned front-loading and included it in the results/discussion sections.

Minor Comments:

1) p.1 l.11 "Reproductions" -> "Reproduction"

We rephrased this sentence, such that it now reads:

ll. 10-12: “NAO flow pattern reproductions in the CanESM2-LE trigger responses in the high-resolution CRCM5-LE that are comparable with reanalysis data.”

2) p.2 l.23 "Atmospheric modes" -> "Such large-scale atmospheric modes"

Thank you, we rephrased accordingly (l.24-25).

3) p.2 ll.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." perhaps reword to, "One way to sample realizations is to perform an initial-value ensemble in which multiple simulations are performed with identical external forcings but perturbed initial conditions. Such an initial-value ensemble would allow a more robust distribution of NAO index values to be sampled."

In this case, we opt for a combination of the suggestion and the original sentence:

ll. 55-62: “One way to sample realizations is to perturb the initial conditions of the model, leading to multiple simulations with identical external forcing which only differ due to internal variability. […] Such initial condition ensembles also allow a more robust distribution of atmospheric modes to be sampled, as was done e.g. for El Niño/Southern Oscillation in Maher et al. (2018).“

4) p3. l.78 "until 2099" -> "that extend until 2099" ***and*** 5) p.4 l.94 "regarding" -> "of"

Thanks, we adopted both wording suggestions (l. 88, l. 104).

6) p.5 l.97 At this point I didn't recall what IMS represented. Even though it was defined on the previous page under key question c, I had to stop and go back to find it. It is only used once in this paragraph and not used again for 5 pages. The economy of saving three words is not worth the break in flow here. This is a problem that exists all over the manuscript and it does significant harm to the readability of the paper. I would encourage the authors to greatly reduce the use of such acronyms - particularly when used so intermittently.

We removed the acronym IMS (as well as REF) which indeed seem superfluous. Several others, such as GCM, RCM, NAO or the models’ and variables names were kept, as we think that they are intuitive to the reader and make the text more readable. When it comes to the region names, we added the word “region” to the acronyms to break long “acronym chains”.

7) p.5 ll.100-104 Care should be taken here. Just because a globally integrated quantity from an initial-value ensemble of one model spans a similar range as a multi-model ensemble does not mean they are interchangeable. Different models have different physics packages and so could have very different regional behavior and that fact is potentially masked by a globally integrated diagnostic.

Thank you for this remark. We rephrased the paragraph regarding this aspect. We did not mean to indicate that both spreads were interchangeable. They are not due to different sources of uncertainties in both ensembles (internal variability vs. model response uncertainty + internal variability). We included a short discussion on the subject of spatially integrated versus regional/local signals:

ll. 108-117: “Comparing the internal variability of the CRCM5 members with the inter-member spread of a subset of the multi-model EURO-CORDEX (Coordinated Regional climate Downscaling Experiment) ensemble regarding regionally integrated European 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 comparison of the single-model and multi-model spreads suggests that a large fraction of the CORDEX ensemble spread regarding temperature and precipitation 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). At smaller regional scales, however, single-model and multi-model spreads may show considerable and in parts temporally changing differences which may partly be induced by model response uncertainties (von Trentini et al., 2019).”

8) p.5 l.118 "Figure 1" -> "In the Appendix, Figure 1"

We included the suggestion: “In the Appendix, Fig. A1, …” (l. 131).

9) p.5 ll.120-124 Perhaps reword "The GCM... PR sum values" to "Displaying opposite bias to CRCM5, the GCM overestimates (underestimates) mean winter nSAT in the norther (southern) part of the domain, whereas winter PR sum is underestimated in the eastern half of the domain and overestimated on the western side of the Alps. As this study will focus on responses in nSAT and PR induced by the NAO (see Section 2.2.4), aside from regions with particularly high PR sum values, it is found that such NAO responses are generally insensitive to these biases."

Thanks for this suggestion. We included it with additional figure references:

ll. 133-138: “Displaying opposite bias to CRCM5, the CanESM2 overestimates (underestimates) mean winter nSAT in the northern (southern) part of the domain (Fig. A1 (b)), whereas winter PR sum is underestimated in the eastern half of the domain and overestimated on the western side of the Alps (Fig. A1 (e)). As this study will focus on responses of nSAT and PR induced by the NAO (see Section 2.2.4), aside from regions with particularly high PR sum values, it is found that such NAO responses are generally insensitive to these biases.”

10) p.5 ll.127-128 "...dispersion. So the following analyses were not confined to..." -> "...variability. So, in addition to analyses of"

We adopted this suggestion (l.141).

11) p.8 ll.199-200 "As a next step, the monthly difference between driving data and the RCM data was taken for each time step and member" -> "As a next step, time series of the difference between monthly mean driving data and the RCM data was taken for each member"

When revising the methods section, we changed this paragraph. In accordance with a comment raised by the second reviewer, we also dropped the asterisk: RMS → RMSD. The paragraph now reads:*

ll.211-213: “As a next step, a root-mean-square difference (RMSD) of the difference time series between monthly mean driving and RCM data over the hist and fut time periods is obtained across all members and winter months:“

12) p.8 l.205 "between driving" -> "between monthly mean driving"

Thanks, we adopted this clarifying suggestion (l. 216).

13) p.8 ll.208-209 "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." -> "provides a measure relative to the inter-annual variability of the SLP pattern in a given location. Low RMS∗ values in a particular region indicate a low error and so a good reproduction of the SLP variations in that region of the RCM."

Thank you, we accepted this suggestion apart from the last sentence. Instead, we included an according explanation in the results section (by following your suggestion in point 21).

ll. 219-220: "The normalization by the square root of the temporal variance of the driving data provides a measure relative to the inter-annual variability of the SLP pattern in a given location."

14) p.9 l.228 "IMS" see my point 6)

We aimed at solving this problem, e.g. by reminding the reader what a given abbreviation means. "IMS" is, as indicated above, removed now.

15) p.9 ll.227-234 This is poorly worded. The authors are mixing ideas of stationary and non-stationary systems to discuss issues related internal variability. For stationary dynamical systems, one can define/identify internal variability by looking a the statistical properties of a time series relative to its long-term time average. For a non-stationary system, one can define/identify internal variability by looking a the statistical properties of an initial-value ensemble of time series relative to its ensemble mean at any specific time. The degrees of freedom that go into each evaluation depends on the length of the time series in the former and the number of ensemble members in the latter.

Thank you for your explanation. We revised this paragraph, including some additional references suggested by the second reviewer.

16) p.10 l.248, l.251 The terms ERA-I data and REF realization are the same I believe. This is very confusing as the equivalence between the two was drawn 12 pages earlier! Why REF? Why not OBS. This would be much more obviously connected to ERA-I. This again falls back to my point 6). Also l.254 I couldn't remember what IMS was again here.

The abbreviations REF, GCM and RCM refer to the models; ERA-I, CanESM2 and CRCM5 are the models'/model data names. We refrained from the abbreviation REF and changed to ERA-I or reference. We did not use OBS since using "reanalyses" and "observations" synonymously may confuse readers.

17) p.10 l.260 l.261 My point 6) again. I had to hunt a long time to find where SOIC, MOIC and "n" were defined. The first two weren't even in the text - they were in the caption to Fig. 1! Again, this breaks the flow of the paper and makes it very difficult to read and understand. Are these acronyms really necessary? For "n" perhaps remind the reader what it is when using it here - even a text search of a PDF isn't much help with a single letter!

SOIC and MOIC are explained in the text in line 260 (first revised manuscript version): "(i) correlations among the 10 members from the same ocean family (same ocean initial conditions in 1950, SOIC " […] "(ii) correlations between each member and the 40 members from the 4 other ocean family (mixed ocean initial conditions, MOIC".

To clarify, we rearranged the order of the words, such that this phrase now reads:

ll. 284-286: "(i) correlations among the 10 members from the same group (same ocean initial conditions – SOIC, n = 225 cases, dotted lines in Fig. 3(b)) and (ii) correlations between each member and the 40 members from the 4 other groups (mixed ocean initial conditions – MOIC, n = 1000 cases, solid lines in Fig. 3 (b))"

n as symbol for sample size was introduced in line 249/250 of the first revised manuscript version: "… due to a larger *sample size (n* *= 1500 for CanESM2-LE and n = 30 for ERA-I)." We followed your suggestion, including a reminder.*

18) p.10 ll.254-260 "For further ... in Leduc et al. (2019):" In addition to comments 16 and 17 above, I found this description very hard to follow. In its place I offer the following. "The independence of the 50-member ensemble is critical to interpreting the inter-member spread as a proxy for internal variability. In evaluating this, it is important to recall that the 50-member LE was constructed in two parts. First, independent atmosphere/ocean states in 1850 were used to launch 5 historical simulations and integrated forward until 1950. Second, in 1950, each of these 5 ensemble members were used to launch 10 individual simulations by applying a small perturbation to the atmosphere and integrated forward until 2099 thereby producing the 50-member large ensemble. As a consequence, for this study, members between each of the 5 groups of 10 are expected to be highly independent while members within each group of 10 are perfectly correlated in 1950 and progressively increase their independence beyond their 1950 starting point. To evaluate whether the 10 members within each of the 5 groups had become sufficiently independent by the two 30-year periods of interest (1981-200 and 2070-2099), correlations were applied to two groups following Leduc et al. (2019):"

Thank you for this detailed suggestion. We adopted this paragraph with some minor adjustments:

ll. 274-284: "For the following analyses independence of the 50-member ensemble is critical to interpreting the inter-member spread as a proxy for internal variability. In evaluating this, it is important to recall that the 50-member CanESM2-LE was constructed in two steps (Fyfe et al., 2017; Leduc et al., 2019): First, independent atmosphere/ocean states in 1850 were used to launch 5 historical simulations integrated forward until 1950. Second, in 1950, each of these 5 ensemble members were used to launch 10 individual simulations by applying a small perturbation to the atmosphere and integrated forward until 2099, thereby producing the 50-member large ensemble.

As a consequence, for this study, members between each of the 5 groups of 10 are expected to be independent. However, members within each group of 10 are highly correlated in 1950 and progressively increase their independence beyond their 1950 starting point. To evaluate whether the 10 members within each of the 5 groups have become sufficiently independent by the two 30-year periods of interest (1981-2010 and 2070-2099), correlations among member time series are applied to two groups following Leduc et al. (2019):"

19) Fig.3 Perhaps plot the normal distribution in this figure for reference.

Thank you, we followed this suggestion (see green line in Fig. 3):

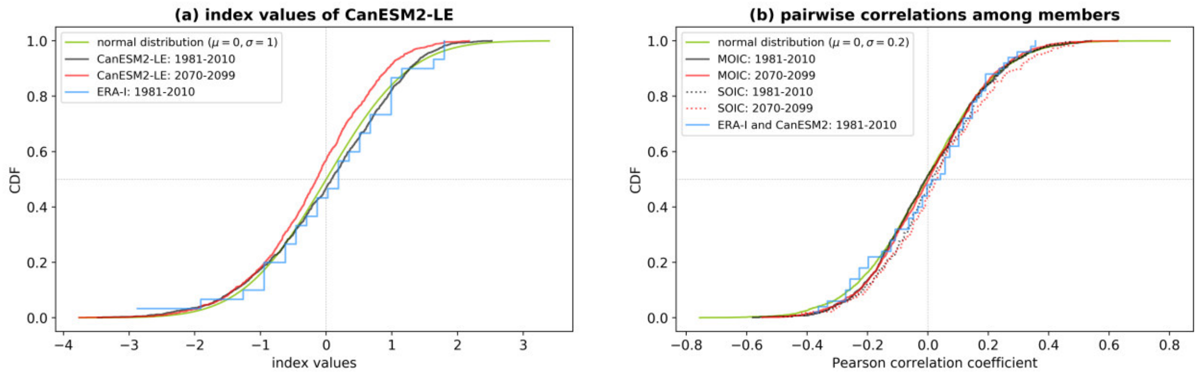


Figure 3: Cumulative density functions (CDFs) of NAO index values. (a) distribution of all CanESM2-LE (n = 50 x 30 per period) and ERA-I (n = 30) NAO index values. (b) pairwise correlations among member NAO index time series from the same ocean families (SOIC – same ocean initial conditions, dotted lines, 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, green: normal distribution with μ = 0 and σ = 1 in (a) and σ = 0.2 in (b).

20) p.12 l. 297 "RMS*" -> "RMS* (eq. 2)" it's been many pages since it was discussed. Which brings up an important point. Why not motivate the need for this diagnostic, then discuss the derivation of RMS* at the place where the diagnostic is used - here. This goes for most of what is presented in Section 2 and then is used in later sections of the manuscript. It would make the paper far more readable. The constant back-and-forth looking for things being discussed is quite disruptive. The authors are making it very challenging for themselves by this choice of presentation and they are not doing a good job of meeting this challenge.

Thanks for your concerns. As mentioned in the beginning of this document, we decided to keep the separation of methods and results. In this case, we included an introducing sentence which may also be seen as a reminder of concept, before explaining the RMSD:

l. 324: "This is achieved by analysing the deviations of RCM and driving data SLP variability."

21) p.12 ll.298-300. "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." To make this point clearer, perhaps reword to, "An O(1) value of RMS* would indicate a poor reproduction of the SLP signal in the RCM because the RMS difference between the RCM and driving data SLP is of the same order as the variability of the SLP in the driving data itself. Values of RMS*<<1, on the other hand, would indicate a good reproduction of the SLP signal in the RCM because it suggests that the RCM is tracking the variability in the driving data."

Thanks, we adopted this suggestion with some minor changes:

ll. 326-329: "An O(1) value of RMSD would indicate a poor reproduction of the SLP signal in the RCM because the RMSD between the RCM and driving data SLP is of the same order as the variability of the SLP in the driving data itself. Values of RMSD << 1, on the other hand, would indicate a good reproduction of the SLP signal in the RCM because it suggests that the RCM is tracking the variability in the driving data."

22) p.12 ll.300-302 With 21) above you could now change, "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..." to "With this understanding it can be seen that the

large-scale SLP pattern over the entire ClimEx domain, which also includes the CEUR, NE, BY and SE domains, is reasonably well represented in most parts of..."

Thanks, we also adopted this suggestion, also with some minor modifications:

ll. 329-331: “With this understanding it can be seen that the large-scale SLP pattern is reasonably well represented in most parts of the entire ClimEx domain for both driving data sets and both periods…”

23) p.13 ll.311-313 "nSAT", "PR", "\alpha_1" see my point 6) and 20). I've run out of steam highlighting these and I leave it to the authors to identify and correct the many that remain beyond this point in the paper

As stated before, we tried to reduce our use of acronyms or abbreviations. However, in this case, since nSAT and PR are the names of our variables, we did not remove them. But we did include a reminder on what “\alpha_1” means.

ll. 340-342: “nSAT and PR spatial responses as revealed in the ERA-I data are generally reproduced under current climate conditions in the CanESM2-LE and CRCM5-LE (see Figs. 6–8). Highest magnitudes of the NAO-responses (i.e., the slope of the regression line, $\alpha 1$, introduced in Eq. (4)) occur in the CRCM5/ERA-I run for all variables.“

24) p.13 l.311 "... the ERA-I data are reproduced in their general properties under current ..." -> "... the ERA-I data are generally reproduced under current ..."

Thank you for this suggestion, we adopted it (l. 340).

25) Figs. 6-7 One quantity is displayed in (a)-(f) and a different quantity is displayed in (g)-(j). The titles on the plots in (c)-(f) are identical to those in (g)-(j). This is confusing. Also, it would be helpful if the mean (columns 1 and 3 in rows 2 and 3) and sd (columns 2 and 4 in rows 2 and 3) plots were placed beside each other to facilitate comparison of identical quantities between the GCM and RCM.

We included an axis label to clarify that the first two rows show total values during 1981-2010, whereas the third row shows the 2070-2099 anomalies. Additionally, the panels were rearranged such that the first and second columns show the ensemble mean values for GCM and RCM while the third and fourth column present the ensemble sd.

References:

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, 2017.

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.

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, 2019.

Maher, N., Matei, D., Milinski, S., and Marotzke, J.: ENSO Change in Climate Projections: Forced Response or Internal Variability?, Geophysical Research Letters, 45, 11,390–11,398, https://doi.org/10.1029/2018GL079764, 2018.

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, 2019.

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 internal climate variability on multi-annual time scales. It remains unclear, ~~though~~however, 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) of a nested regional climate model are analyzed for a 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) and the nested regional climate model (RCM; the CRCM5). Responses of mean surface air temperature and total precipitation to changes in the NAO index value are ~~expressed for~~ examined in 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. Results show that statistically robust NAO patterns are found in the CanESM2-LE under current forcing conditions. ~~Reproductions of the NAO flow patterns~~ NAO flow pattern reproductions in the CanESM2-LE trigger responses in the high-resolution CRCM5-LE that are comparable with ~~reference~~ reanalysis data. NAO–response relationships weaken in the future period, but their inter-member spread shows no significant change. The results stress the ~~importance~~ value of single-model ensembles for the evaluation of internal variability by pointing out the large differences of NAO–response relationships among individual members. They also strengthen the validity of the nested ensemble for further impact modelling using RCM data only, since important large-scale teleconnections present in the driving ~~GCM~~ data 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 climate 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), i.e. from varying sequences of weather events 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). ~~Atmospheric~~ Such large-scale 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 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~~ with 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~~ The NAO is commonly quantified with an index 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 station measurements, spatially averaged values of pre-set regions, or the region of highest 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. 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 (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~~, however, lead to highly similar index time series (see e.g., Pokorná and Huth, 2015, for a detailed survey of various approaches).

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: 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). Similarly, the findings regarding the prevalence of future positive or negative states lack unity: Some analyses of CMIP5 models, for example, suggest more positive phases under rising greenhouse gas concentrations until 2100 (e.g., Kirtman et al., 2013; Christensen et al., 2013), others favour an increase of negative phases (Cattiaux et al., 2013).

In most of these studies it was common to rely on one simulation per model and estimate 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 robustly evaluate the range of NAO index values and evolution in a projected future climate, or whether the chosen simulation is a good representation of how this model simulates the phenomenon in question (Leduc et al., 2019). Relying on single realizations possibly deteriorates the assessment of a given model, as single realizations may vary considerably among themselves due to internal variability (and also deviate from the climate evolution observed in reality). One way to ~~trigger~~

~~internal variability in GCM simulations~~ sample realizations is to perturb the initial conditions of the model, leading to ~~several realizations of weather sequences under~~ multiple simulations with identical external forcing which ~~also allow to derive a~~ only differ due to internal variability. Examples for recent GCM initial-condition large ensembles of transient simulations are the 100-member Max Planck Institute Grand Ensemble (MPI-GE, Maher et al., 2019) , the 50-member Canadian Earth System Model Large Ensemble (CanESM2-LE, e.g. Kirchmeier-Young et al., 2017; Fyfe et al., 2017) or the 40-member Community Earth System Modelling Large Ensemble (CESM-LE Kay et al., 2015) which were, among others, used for various analyses of internal variability or extreme events. Such initial condition ensembles would also allow a more robust distribution of ~~NAO index values.~~atmospheric modes to be sampled, as was done e.g. for El Niño/Southern Oscillation in Maher et al. (2018) . That is why ~~this~~ the present 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 characteristics which may be highly relevant in climate impact studies over heterogeneous landscapes (Leduc et al., 2019). Thus, dynamical downscaling of the GCM members using a regional climate model (RCM) is advised (Leduc et al., 2019). ~~Such~~ The downscaling of a GCM single-model large ensemble, the CanESM2-LE, was performed within the Climate Change and Hydrological Extremes project (ClimEx, www.climex-project.org, Leduc et al., 2019).

~~The combination of~~ Examples of analyses on the separation of the forced signal from internal variability within a single-model initial condition GCM-RCM combination of EC-EARTH and RACMO2, though with 16 members, were performed by Aalbers et al. (2018) various extreme precipitation indices.

Combining the driving GCM and nested RCM (i.e., driven by lateral boundary conditions of the GCM) large ensembles (LE) allows for analyzing the spread of NAO states and responses within one model chain~~, thus establishing~~ . In doing so, it is possible to establish the range of internal variability of the NAO ~~, and finding~~ and find robust NAO and response patterns by significantly reducing uncertainty associated with internal variability in the ensemble.

~~This study also targets the~~ The present study targets the research 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~~ GCM-LE and the driven ~~RCM. This issue~~ high resolution RCM-LE. The latter may be important for impact modellers who work with RCM data on internal variability 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 and responses, expressed by the inter-member spread ~~(IMS)~~ among the 50 members of the GCM-LE and the RCM-LE?

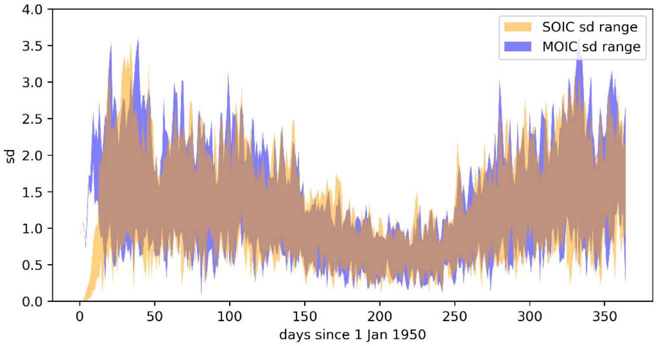


Figure 1. Inter-member standard deviation of a daily NAO index in the CanESM2-LE starting on 1 Jan 1950 as a function of time. The inter-member standard deviation ~~was~~ (sd) is derived from ~~ten~~ 10 groups of ~~five~~ 5 members with the same ocean initial conditions (SOIC) and ~~ten~~ 10 groups of ~~five~~ 5 members with mixed ocean initial conditions (MOIC, following an approach in Leduc et al., 2019).

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

2 Data and Methodology

2.1 Data

Data from three different sources ~~were~~ are employed in this study (Table 1). The major source ~~was the LE~~ is the RCM-LE data set of the ClimEx project which is described in detail in Leduc et al. (2019). The ClimEx project 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). 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~~ 10 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 (see Leduc et al., 2019, for a similar presentation of member independence).

As described in Leduc et al. (2019), ~~these~~ the 50 ~~GCM~~ CanESM2 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~~. During the nesting process, large-scale spectral nudging regarding~~ , each sized 280×280 grid cells on a rotated grid. Large-scale spectral nudging of the horizontal wind field was applied during the nesting process (Leduc et al., 2019). This single-RCM 50-member ensemble allows for internal variability and extreme events to be detected in high

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

CCCma – Canadian Centre for Climate Modelling and Analysis

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 ~~IMS~~ inter-member spread of a subset of the multi-model EURO-CORDEX (Coordinated Regional climate Downscaling Experiment) ensemble regarding regionally integrated European 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~~ comparison of the single-model and multi-model spreads suggests that a large fraction of the CORDEX ensemble spread regarding temperature and precipitation 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). ~~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~~ At smaller regional scales, however, single-model and multi-model spreads may show considerable and in parts temporally changing differences which may partly be induced by model response uncertainties (von Trentini et al., 2019) .

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

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

- (mean) sea level air pressure (referred to as “SLP”, converted to $[\mathrm{hPa}]$) to obtain the NAO,
- near surface air temperature (referred to as “nSAT”, converted to $[\mathrm{K}]$),
- total precipitation including liquid and solid precipitation from all types of clouds (referred to as “PR”, converted to $[\mathrm{mm}]$).

ERA-I variables t2m, tp and msl ~~were~~ (Table 1) are chosen as they ~~were~~ are assumed to most accurately represent the ~~variables from the~~ GCM and RCM ~~models~~variables. As the variables derived from the three data sources ~~were originally available with~~ are available at different temporal resolutions (three-hourly for tas in RCM, hourly for pr in RCM, daily ~~in GCM~~ for psl, pr and tas in GCM, 6-hourly for ERA-I t2m and msl analysis, and 12-hourly for ERA-I tp forecast data), they ~~were~~ are all aggregated to daily values first.

~~Figure A2~~ In the Appendix, Fig. 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~~ Fig. A1 (a) for ERA-I and ~~third column~~ (c) for CanESM2), whereas winter PR sums are overestimated in nearly the entire domain with strongest values in the south-eastern part ~~. The GCM~~ (Fig. A1 (d) and (f)). Displaying opposite bias to CRCM5, the CanESM2 overestimates (underestimates) ~~nSAT mean north (south) of the Alps.~~ mean winter nSAT in the northern (southern) part of the domain (Fig. A1 (b)), whereas winter PR sum is underestimated in the ~~entire domain apart from~~ eastern half of the domain and overestimated on the western side of the Alps ~~in the GCM. However, as~~ (Fig. A1 (e)). As this study will focus on ~~changes in~~ responses of 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~~ aside from regions with particularly high PR sum values, it is found that such NAO responses are generally insensitive to these biases.

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 to not only influence winter mean values, but also their ~~dispersion. Sothe following analyses were not confined to~~ inter-annual variability. So, in addition to analyses of winter mean temperature (nSAT mean) and precipitation sums (PR sum), selected analyses ~~were~~ are also performed on winter mean monthly standard deviations of daily mean temperature (nSAT ~~std~~sd) as a measure of temperature variation.

2.2 Methodology

2.2.1 Regions of interest and time horizon selection

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

The regions of interest and their names used in this study are displayed in Fig. ~~1~~2. The formation of the NAO over the North Atlantic (NAR, AH, IL regions~~, see annotations in Fig. 1) was~~) is analyzed in the ERA-I and CanESM2-LE dataset, while responses over Central Europe (CEUR, NE, BY, SE ~~) were~~ regions) are evaluated in ERA-I, CRCM5/ERA-I, CRCM5-LE and CanESM2-LE data.

AH and IL regions are 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 SLP centres (see e.g., Moore et al., 2013), both NAO core regions ~~were~~ are extended to 3×3 GCM grid cell matrices. ~~The NAO index proved~~ In preliminary analyses conducted for the present study, the NAO index has proven to be very robust towards the exact shape of the core regions~~in preliminary analyses~~.

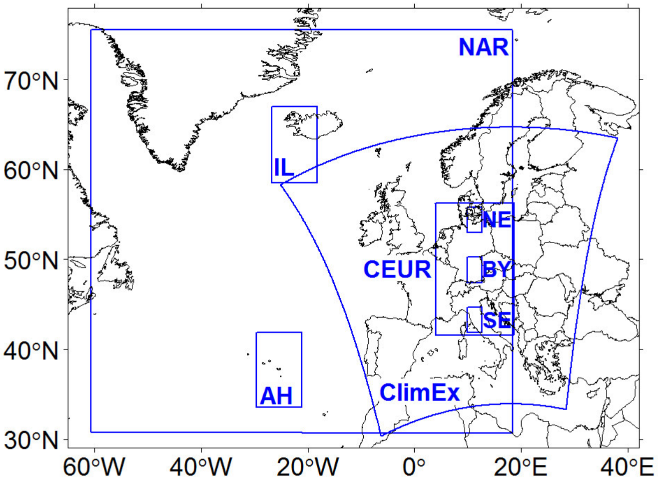


Figure 2. Regions of interest. Abbreviations and domain sizes in terms of GCM grid cells (2.8°) are as follows: AH – Azores High (3 × 3); IL – Icelandic Low (3 × 3); NAR – large-scale North Atlantic region (28 × 16); CEUR – Central Europe (5 × 5); NE – northern Europe (1); BY – Bavaria (1); SE – southern Europe (1); ClimEx – domain used in ClimEx project (extent approximately 22 × 12 after resampling to GCM grid).

The Central European domain (CEUR) ~~was~~ is defined in the CanESM2-LE by selecting a 5 × 5 GCM grid cell matrix 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 within the ClimEx European domain ~~was~~ is 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 the NAO ~~were~~ are expected to vary over the CEUR domain, it ~~seemed~~ seems 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) denote small-scale 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~~ is 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~~ is assigned to it. ~~REF~~ ERA-I and RCM data (3 × 4 and 26 × 26 grid cells, respectively) ~~was~~ is spatially aggregated to GCM resolution for this part of the analysis.

This study ~~focused~~ focuses on inter-annual analyses which ~~were~~ are conducted for two time horizons covering 30 years each. The historical (hist; 1981–2010) period is used to establish reference statistics in the ERA-I data and the ERA-I driven CRCM5 run which are then evaluated in the GCM-LE and the RCM-LE. Links and relationships established for the historical period are also investigated in a far future horizon (fut; 2070–2099). The chosen period length ~~was~~ is 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.

Relationships between the NAO and 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 multi-decadal (>30 years) NAO–response variability (Woollings et al., 2015), stationarity in NAO-impact relationships ~~was~~ is assumed for simplicity reasons. ~~The historical (hist; 1981–2010) period was used to establish reference statistics 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 nSAT and winter sums for 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~~ are performed for this season only. Preliminary tests ~~had~~ within this study have shown that correlations and links between the NAO index and the climate variables ~~were~~ are more distinct from noise, if March ~~was~~ is included as well. That is why an extended winter season ~~was~~ is used here (DJFM, see also Iles and Hegerl, 2017; Hurrell, 1995; Osborn, 2004).

All data (spatially explicit and subset time series) is aggregated to the seasonal time scale for further use (winter means for nSAT and winter sums for PR).

2.2.2 Deriving an NAO index

The NAO index ~~was~~ is derived from ERA-I and CanESM2-LE data, resulting in 1 ~~REF~~ ERA-I and 50 GCM realizations. As the CRCM5 ClimEx domain does not cover the AH and IL regions (see Fig. 2), the index is not derived from this data source. 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~~ It therefore directly represents the winter SLP gradient over the North Atlantic.

The time series of AH and IL ~~originated~~ originate from the temporally shortened and spatially averaged SLP time series of both grid cell matrices~~for REF and GCM data only. As the CRCM5 ClimEx domain does not cover the AH and IL regions (see Fig. 1), the index was not derived from this data source.~~. Daily SLP values ~~were~~ are averaged to monthly means (Cropper et al., 2015) and scaled to obtain 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}):

$$\text{NAOIndex} = \frac{AH - \overline{AH}}{s_{AH}} - \frac{IL - \overline{IL}}{s_{IL}} \tag{1}$$

Monthly indices ~~were~~ are next averaged to DJFM means. This approach is similar to Woollings et al. (2015) and Jones et al. (2013).

The ERA-I NAO index caluclated this way shows high agreement with often cited NAO indices like the time series of Hurrell (Pearson correlation of $r = 0.95$ with ERA-I NAO index; index available at https://climatedataguide.ucar.edu/climate-data/hurrellnorth-atla For further analyses it will therefore serve as a reference.

To compare future with historical index values, the future time series of AH and IL ~~were normalized~~ are standardized with the present SLP standard deviations (see also Ulbrich and Christoph, 1999; Hansen et al., 2017) and mean values. The ~~normalization~~ standardization of each GCM member ~~was~~ is carried out individually~~, such that each member has specific normalization parameters~~.

2.2.3 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~~ is verified. Therefore, monthly mean SLP data of the CRCM5 (both driving data sets) and ERA-I ~~were~~ are linearly interpolated to GCM resolution over the ClimEx domain. During interpolation, small scales ~~were~~ are 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~~ a root-mean-square difference (RMSD) of the difference time series between monthly mean driving and RCM data over the hist and fut time periods ~~was obtained which was later averaged over all ensemble members and the~~ is obtained across all members and winter months:

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

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

where $\langle\cdot\rangle$ is the averaging operator over a given index, D_m is the difference between monthly mean driving data and RCM data; Drive_m is driving SLP data; $\mathrm{VarDrive}_m$ is the variance of ~~the~~ SLP driving data over the 30 year periods; i,j are spatial ~~coordinates of the grid~~ grid coordinates, m are months 12, 1–3, n are ensemble members 1–50 ~~and~~ for CanESM2 and 1 for ERA-I, 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~~ provides a measure relative to the inter-annual variability of the SLP pattern ~~on~~ in a given location. ~~Low RMS* values indicate a low error.~~

2.2.4 Climatic Changes Associated with NAO

All data sources (Table 1) ~~were~~ are used to obtain response patterns of the variables nSAT and PR. Climatic changes associated with the NAO ~~were~~ are evaluated using Pearson correlation coefficients and a slope parameter obtained by linear regression. ERA-I and CRCM5/ERA-I nSAT and PR data ~~were~~ are correlated with the ERA-I index, CanESM2 and CRCM5 members ~~were~~ are correlated with the CanESM2 index calculated for the corresponding member.

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

$$Y = \alpha_1 X + \alpha_0 + \varepsilon_Y \qquad (4)$$

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~~ is estimated on each grid cell using ordinary least squares regression with the R function lm (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. (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~~ is preferred in order to take into account the member-specific index units. The NAO–response relationship is analyzed individually for each GCM and RCM member (as is done e.g. in Woollings et al., 2015) .

2.2.5 Addressing Internal Variability

~~Internal variability was understood as being~~ In this study, the GCM-RCM combination allows to set a focus on the internal variability of an RCM ensemble and the driving GCM ensemble. Climate modes tend to show high internal variability (see e.g. Maher et al., 2018, for an analysis of ENSO internal variability in CMIP5 models and two single-model large ensembles) . The present study targets the NAO-related internal variability within one GCM-RCM combination.

In general, natural internal variability may be understood from different angles. When looking into single realizations of time series of a given variable, internal variability may be seen as represented by the ~~oscillations~~ oscillation around the long-term mean ~~of the time series of a given variable (Hawkins and Sutton, 2011)~~ evolution, i.e. the residuals (Frankcombe et al., 2015; Hawkins and S 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~~ case, the amplitude of internal variability is usually calculated as a time-invariant quantity during the period under consideration (Hawkins and Sutton, 2009, 2011) . Another way is investigating transient internal variability in initial-condition ensembles, like e.g. in Maher et al. (2019) . In this case, the ensemble establishes ranges of possible weather event sequences by superposing single realizations which are 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.~~ .

In the present study, the latter approach is used within the 50-members CanESM2-LE and CRCM5-LE. This allows to sample internal variability at single points in time as the range of the members' values, i.e. across members (e.g., Maher et al., 2018) . While internal variability is assumed to be stationary within both 30-year periods for this study, the use of a LE allows to detect potential changes in internal variability between both analysis periods.

~~Therefore, the NAO–response relationship was analyzed individually for each GCM and RCM member (as is done e.g. in Woollings et al., 2~~ ~~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~~ Internal variability is expressed as the across-member standard deviation, i.e. the inter-member spread of the CanESM2-LE and the CRCM5-LE (see also Leduc et al., 2019; Déqué et al., 2007; Aalbers et al., 2018) among the 30-year means, rather than computing a transient internal variability at each time step as was done e.g. in Maher et al. (2019) . Aggregations to ensemble means (like in Deser et al., 2017; Aalbers et al., 2018) of NAO responses are only performed for illustrating purposes in order to avoid 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

Naturally, the first step when evaluating the NAO in a model ensemble is to analyse its representation and index distribution in the model data of interest.

3.1.1 NAO index and SLP conditions

~~First, a 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). For further analyses it will therefore serve as a reference. Cumulative density functions (CDFs) of NAO index values. (a) distribution of all CanESM2-LE ($n = 50 \times 30$ 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 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. 2~~ similar to a normal distribution with $\mu = 0, \sigma = 1$, Fig. 3 (a))~~. The~~ , but 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. ~~2~~ 3 (a)) of some of the 50 members exceed those of the ~~REF~~ ERA-I realization; thus, ~~the REF~~ ERA-I which serves as a reference realization lies well within the ensemble ~~IMS~~inter-member spread. The future NAO index shows a similar distribution of values, but with slightly less positive and more negative values (red curve in Fig. ~~2~~ 3 (a)). ~~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 members would contradict the assumption of independence. In order to take into account the two perturbations during~~

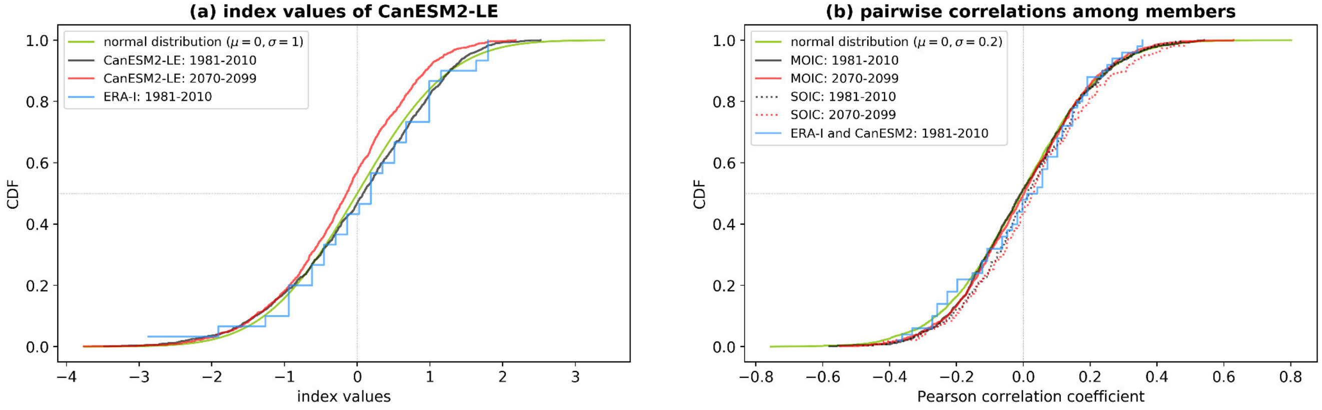


Figure 3. Cumulative density functions (CDFs) of NAO index values. (a) distribution of all CanESM2-LE ($n = 50 \times 30$ per period) and ERA-I ($n = 30$) NAO index values. (b) pairwise correlations among member NAO index time series from the same ocean families (SOIC – same ocean initial conditions, 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. Green: normal distribution with $\mu = 0$ and $\sigma = 1$ in (a) and $\sigma = 0.2$ in (b).

~~the production of the LE (~~

For the following analyses independence of the 50-member ensemble is critical to interpreting the inter-member spread as a proxy for internal variability. In evaluating this, it is important to recall that the 50-member CanESM2-LE was constructed in two steps (Fyfe et al., 2017; Leduc et al., 2019) : First, atmospheric perturbations in 1850 ~~for~~ were used to launch 5 ~~ocean families,~~ historical simulations integrated forward until 1950. Second, in 1950~~for perturbations leading to~~ , each of these 5 ensemble members were used to launch 10 ~~members per ocean family), these correlations were split in two groups like in~~ individual simulations by applying a small perturbation to the atmosphere and integrated forward until 2099, thereby producing the 50-member large ensemble.

As a consequence, for this study, members between each of the 5 groups of 10 are expected to be highly independent. However, members within each group of 10 are perfectly correlated in 1950 and progressively increase their independence beyond their 1950 starting point. To evaluate whether the 10 members within each of the 5 groups have become sufficiently independent by the two 30-year periods of interest (1981-2010 and 2070-2099), correlations among member time series are applied to two groups following Leduc et al. (2019): (i) correlations among the 10 members from the same ~~ocean family~~ group (same ocean initial conditions ~~in 1950,~~ – SOIC, $n = 225$, ~~see~~ dotted lines in Fig. ~~2~~3 (b)) and (ii) correlations between each member and the 40 members from the 4 other ~~ocean families~~ groups (mixed ocean initial conditions ~~,~~ – MOIC, $n = 1000$, ~~see~~ solid lines in Fig. ~~2~~ 3 (b)).

These correlations approximately follow a normal distribution with $\mu = 0$ and $\sigma = 0.2$. There is a slight surmount of low positive correlations in the SOIC group compared ~~to~~ with the MOIC group which is (not significantly) stronger in the fut time horizon (see red and black dotted lines in Fig. ~~2~~ 3 (b)). Although zero correlations do not necessarily imply independence, clear correlations among members would contradict the assumption of independence. 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~~ ERA-I NAO index despite similar statistics (see also Fig. ~~8~~9). Thus, the ERA-I and GCM indices can be seen as not dependent realizations drawn from the same distribution.

In ~~addition to the NAO index evaluation~~order to further evaluate the NAO representation in the CanESM2-LE, Fig. ~~3~~ 4 presents the large-scale SLP patterns in the NAR region during neutral, positive and negative NAO conditions. Positive (negative) index years are chosen, if the respective index value exceeds 1 (-1) as in Rogers (1984). The neutral conditions refer to the 30 year SLP average. Regions with strong sampling uncertainties, i.e. where the standard error is larger than the anomaly, are indicated with stippling in panels (b)–(c) and (e)–(f).

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. 3)compared to REF (top row)~~Fig. 4 (d)) compared with ERA-I (Fig. 4 (a)). The mean SLP difference between the CanESM2-LE mean and ~~REF~~ ERA-I 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~~ with ERA-I (~~panels~~ Fig. 4 (a), (d)~~in Fig. 3~~). Long-term neutral states of both driving data sources show robust signals in the entire NAR region (i.e., no stippling). This suggests that the different patterns in GCM and ~~REF~~ reference data are not singularly artefacts arising from different sample sizes~~.~~, but rather robust features.

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~~reference data (Fig. 4 (b)–(c) and (e)–(f)~~. These regions are mostly found in the transition region between the wider ERA-I~~). Transition regions between the AH and IL nodes are marked by high uncertainty in ERA-I, 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 in Fig. 4 (e)–(f). So the difference between CanESM2-LE and ERA-I NAO anomalies may be due to the fact that ERA-I composites are derived from 3 negative and 4 positive years whereas the GCM data provides 264 negative and 263 positive years during 1981–2010.

The difference between SLP anomalies in positive and negative years representing the pressure variability is indicated by white lines. ~~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~~ERA-I) 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 SLP values in the neutral state, but are situated near the 3×3 GCM grid cell matrices used for index calculation. This ~~is very promising as it~~

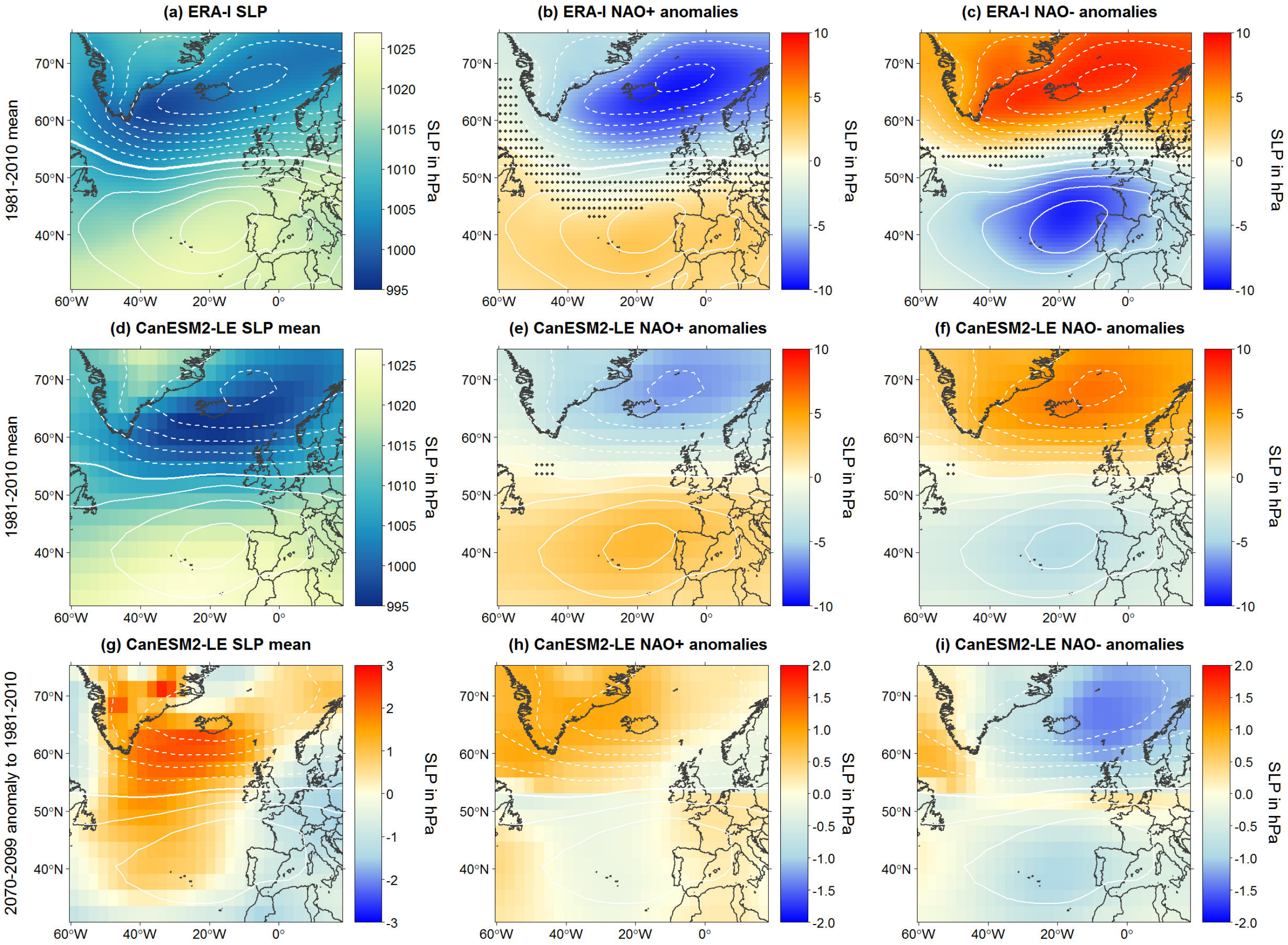


Figure 4. NAR winter mean SLP [hPa] composites in ~~REF~~ ERA-I ((a)–(c)) and ~~GCM~~ CanESM2-LE ((d)–(i)) data showing long-term neutral conditions (left column), NAO positive (mid column) and negative anomalies (right column). (a)–(f): for 1981–2010. (g)–(i): 2070–2099 changes with respect to 1981–2010 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~~ Grey stippling in subpanels (~~a~~b)–(c) and (e)–(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).~~

supports the choice of these SLP centres for index calculation.

Under projected future climate conditions, 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 SLP shows a marked increase in the northern NAO node region. Negative phases exhibit SLP decreases in both node regions, although with larger changes near IL, resulting in negative phases to become slightly weaker as well.

Having established a reasonably plausible representation of the NAO in the driving data, the next step is to evaluate the large-

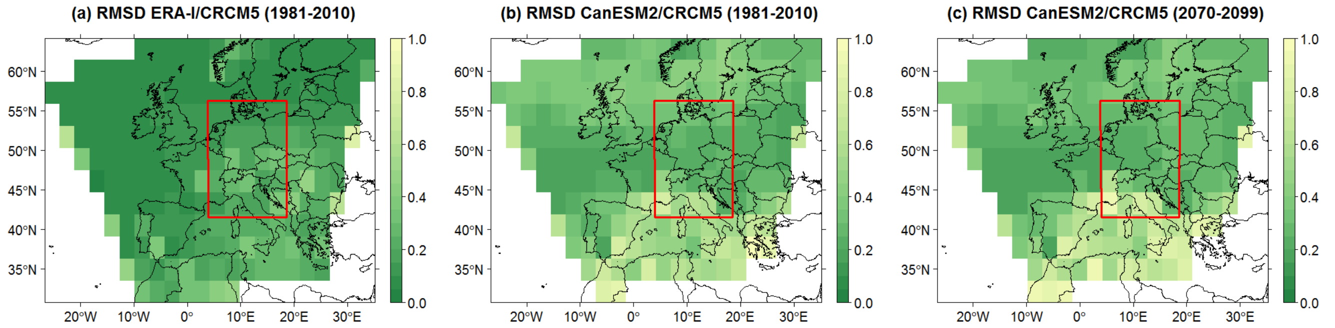


Figure 5. ~~RMS*~~ RMSD of monthly SLP differences between driving data and CRCM5 members, calculated following Eq. (2). Colouring: ~~RMS*~~ RMSD ≤ 1 significant at $p \leq 0.05$ with a false detection rate smaller than 0.1 (see Wilks, 2016). (a) for driving data ~~set~~ ERA-I (1981–2010, one realization), (b) for driving data ~~set CanESM2~~ CanESM2-LE (1981–2010, 50 members), (c) for driving data ~~set CanESM2~~ CanESM2-LE (2070–2099, 50 members). Red box: position of CEUR domain.

scale NAO pattern in the RCM data. ~~Figure 4 maps the RMS*~~ This is achieved by analysing the deviations of RCM and driving data SLP variability. Figure 5 maps the RMSD of the difference between driving data and RCM SLP during 1981–2010 for driving data ERA-I (a) and ~~CanESM2~~ CanESM2-LE (b)~~and~~ , and CanESM2-LE in 2070–2099 (c). ~~A value of RMS* ≥ 1 indicates that the root-mean squared error~~ An $O(1)$ value of RMSD would indicate a poor reproduction of the SLP signal in the RCM because the RMSD between the RCM and driving data ~~is larger than the temporal variability~~ SLP is of the same order as the variability of the SLP in the driving data ~~. In this case, the large-scale SLP pattern may not be seen as being correctly represented~~ itself. Values of RMSD $<< 1$, on the other hand, would indicate a good reproduction of the SLP signal in the RCM ~~data. The~~ because it suggests that the RCM is tracking the variability in the driving data. With this understanding it can be seen that 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 $p \leq 0.05$ 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*~~ subpanels in Fig. 5 show an RMSD 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*~~ RMSD 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. 4~~ Fig. 5 (c)). The differences of the spatial patterns are most likely due to different large-scale SLP patterns in both driving data sets. In the CEUR domain (~~indicated as~~ red box in Fig. ~~4~~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~~

nSAT and PR spatial responses as revealed in the ERA-I data are ~~reproduced in their general properties~~ generally reproduced under current climate conditions in the CanESM2-LE and CRCM5-LE (see Figs. ~~5~~6–~~7). Responses with the highest~~ 8). Highest magnitudes of the NAO-responses (i.e., the slope of the regression line, α_1 ~~magnitudes~~ , introduced in Eq. (4)) occur in the CRCM5/ERA-I run for all variables. In general, the CRCM5 produces stronger α_1 response values at the local scale than the driving data. Regarding the absolute α_1 values, the CRCM5-LE ~~values meet~~ mean meets 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. 5~~Fig. 6) and less day-to-day nSAT variability ~~(see Fig. 6~~compared to neutral conditions (Fig. 7). The generally positive relationship between nSAT mean and NAO (~~see Fig. 5~~Fig. 6) is strongest in the north-eastern parts of the domain. Regionally, the NAO explains up to 40–60 % of nSAT mean variability (see also Fig. ~~A3~~ 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 nSAT variability reaches up to 0.4–0.6 K in the northeastern continental section while it is near zero in the southern part of the domain.

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. 7~~Fig. 8). The strength of the NAO–PR relationship, expressed by a Pearson correlation coefficient 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. ~~7~~8, panels (a)–(~~c)and (e)~~d)). 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~~ with ERA-I, CRCM5/ERA-I and CRCM5-LE. As is visible in Fig. ~~7~~8, higher α_1 values in mountainous regions indicate strong NAO responses related to orography. Regionally, the NAO accounts for 40–50 % of total 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 sum α_1 values (up to ±220 mm per unit index change)~~where~~. In these parts the 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. ~~A2~~ A1 (f)). PR sum shows only weak correlations in the central region of the CEUR domain.

The mean state of nSAT and PR changes in the transient climate simulation towards warmer and moister conditions with less intra-seasonal variability of 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 NAO–climate relationships weaken in general compared ~~to~~ with the historical ones for all variables ~~.~~ as can be inferred from the ensemble mean changes in Figs. 6–8 (g)–(h). The spatial patterns of NAO-induced change do not change considerably between both periods. The response to the NAO, α_1, is clearly reduced in nSAT mean as ~~is nSAT std~~well as nSAT sd, and there is also a reduction in PR sum change (panels (g)~~, (i~~–(h) in Figs. ~~5~~6–~~7~~8).

3.2 Internal Variability at the GCM and RCM scale

The next section now focuses less on the ensemble mean changes, but rather their internal variability. 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~~ the inter-member spreads of the CRCM5-LE ~~compared to~~ and the CanESM2-LE, and their differences.

3.2.1 Multi-member ensemble

The CanESM2-LE reproduces typical NAO index characteristics: Fig. ~~8~~ 9 summarizes several statistics for all 50 GCM members as multiples of the ~~REF~~ reference, i.e. ERA-I, value. Generally, the ensemble meets the ~~REF~~ ERA-I value in all aspects of the NAO index. However, some GCM members only reach half of the ~~REF~~ ERA-I teleconnectivity values (minimum correlation between AH/IL time series: $r = -0.281$, not significantly different from zero at $p \leq 0.05$ using a t-test; ~~REF~~ ERA-I $r = -0.699$). This finding is especially interesting as this metric quantifies the strength of the NAO within the individual members. ~~The IMS~~ The inter-member spread of the teleconnection strength ~~, though,~~ does not change significantly over time, in spite of the SLP changes over the North Atlantic. The 2070–2099 NAO index exhibits less inter-annual variability, less positive phases, more 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 variabilityin the form of~~ NAO-responses also show a considerable degree of internal variability. Its spatial distribution expressed by diverging ensemble members can be derived from Figs. ~~5~~6–~~7 (subplots (d),~~ 8 (e)–(f) ~~)~~ presenting spatially distributed ensemble ~~sd~~ standard deviation (sd) as a measure ~~for IMS.~~ of inter-member spread. Locally, the RCM shows considerably higher spreads than the GCM. Largest deviations for nSAT mean are found in continental regions of CEUR, but they do not simply correspond to high or low α_1 (see also Fig. ~~A4~~ A3 (a)–(d)). ~~Low IMS~~ Low inter-member spread corresponds mostly to Alpine and sea regions. ~~For nSAT mean ,~~ The stippling in Figs. 6–8 (c)–(d) and (g)–(h) indicates regions where the variability among the members is larger than the ensemble mean response, i.e. where the signal-to-noise ~~ratios~~ ratio (SNR) between ensemble mean and sd ~~exceed~~ lies below 1. For nSAT mean, the SNR exceeds 1 in most regions north of the Alps (~~see regions without stippling in Fig. 5)~~Fig. ~~nSAT std~~ 6 (c)–(d)). nSAT sd shows SNR<1 in the northern parts of the CanESM2-LE data (~~see Fig. 6~~ Fig. 7 (c)) and in the Alpine region of the CRCM5-LE data (Fig. ~~6 (e~~7 (d)). This variable shows a strong linear relationship between LE mean and sd (Fig. ~~A4~~ A3 (e)–(h)). Regarding PR sum, RCM members vary most in regions with highest absolute α_1 values and altitudes, but there is no clear dependence in GCM (Fig. ~~A4~~ A3 (i)–(l)). For PR sum, there is an east-west corridor of SNR values below 1 which accompanies rather low α_1 values (~~see Fig. 7)~~Fig. 8 (c)–(d), (g)–(h)).
In addition to future changes in the NAO responses ensemble means, there is also a change in the spatial distribution of the ~~IMS~~ inter-member spread expressed as ensemble sd (~~see subpanels (h),~~ subpanels (i)–(j) in Figs. ~~5~~6–~~7~~8).
To further investigate the ~~IMS~~inter-member spread, Fig. ~~9~~ 10 illustrates the Pearson correlation coefficients r between the NAO index and subset regions for nSAT mean or PR sum in GCM and RCM LEs separately. ~~Both ensemble IMS~~ In these boxplots,

the variability among the members, is illustrated by the boxsize, i.e. the inter-quartile distance. Both ensemble inter-member spreads generally envelope the ~~REF~~ ERA-I value (dashed line) of the given region, apart from GCM hist in Fig. ~~9~~ 10 (b). ~~This~~ This general finding does not change in the projected future climate: most boxes and whiskers keep their size, ~~though GCM PR SE~~ only GCM nSAT in the NE region is characterized by a ~~smaller~~ larger 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). ~~Ensemble~~ Some of the ensemble mean values exhibit a ~~(sometimes significant)~~ significant shift towards lower r values in the future for both models ~~in some regions~~ for nSAT mean and PR sum (see text insertions CanESM2(hist, fut) and CRCM5(hist, fut)). An unpaired Mann-Whitney/U-test ~~was~~ is applied here as the samples from hist and fut ~~were~~ are seen as being drawn from different climates (~~using a χ^2-test,~~ since the null hypothesis of independence between hist and fut periods could not be rejected at $p \leq 0.05$ using a χ^2-test).

3.2.2 Change of scales

Having analyzed GCM and RCM separately so far, ~~it is now advised~~ the next step is to compare both ensembles. A χ^2-test ~~revealed~~ reveals 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~~ the ERA-I reference 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. ~~9~~ 10 for r ~~, where hist(CanESM2, CRCM5) or fut(CanESM2, CRCM5) is indicated; but also Figs. 5~~and Figs. 6–~~7~~ 8 for α_1). This enhancement by the CRCM5 is notably independent of the driving data: for both variables, the CRCM5/ERA-I r value (dotted lines in Fig. ~~9~~10) is also found to be higher than the ERA-I value in most regions (dashed lines in Fig. ~~9).~~ 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 ~~9~~ 10 shows that mean r values of RCM (grey filling) and GCM (~~hollow~~white) members are significantly different in all subset regions for nSAT mean in both time horizons, but only in the NE and BY regions for PR sum; in SE, only weak differences between GCM and RCM PR sum r distributions are visible. In NE and BY regions 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 nSAT mean). 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~~ inter-member spread of the correlation between NAO and response variables is not generally altered during the nesting process.

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~~ are correlated member-wise (see Fyfe et al., 2017, for a similar approach). As can be seen in Fig. ~~10~~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~~ (Fig. 11 (b)) and nSAT sd (Fig. 11 (c)) than for nSAT mean (Fig. 11 (a)) in both periods. Also, the ~~IMS~~ inter-member spread is larger for PR sum and nSAT ~~std~~ sd than for nSAT mean. This finding suggests that there is a larger discrepancy in portraying PR sum and nSAT ~~std~~ sd in the RCM with respect to the GCM compared ~~to nSAT mean.~~ with nSAT mean, i.e. the RCM does not generally track the variability induced by the GCM for these variables. The correlations between CanESM2 and CRCM5 subset regions are in general significantly lower under future climate conditions compared ~~to~~ with the historical ones, apart from nSAT mean in ~~BY~~the BY region, PR sum in ~~SE and nSAT std in BY~~ the SE region and nSAT sd in the BY region (see text in Fig. ~~10~~11). For nSAT ~~std~~ sd a shift of the distribution of r towards slightly larger values is visible. All variables exhibit a future ~~IMS increase, though~~ inter-member spread increase, but not all subset regions are affected (~~see~~ e.g.~~nSAT mean~~ , nSAT mean in BY or nSAT ~~std~~ sd in SE in Fig. ~~10~~11). This suggests that under future climate conditions a ~~considerable~~ potential reduction of GCM–RCM co-variability needs to be ~~taken into account~~considered, 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~~central Europe. Ensemble mean information aggregates several realizations and ~~thus~~ so differences towards the single ~~REF realization may occur~~ERA-I realization are to be expected. However, results ~~showed that the REF~~ show that the ERA-I pattern may in general be seen as being “embedded” in the RCM or GCM ~~IMS~~inter-member spread, implying that GCM, RCM and ~~REF~~ the reference data 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~~ sd 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 NAO.

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. Similar model biases are widely reported (see e.g., Ruprich-Robert and Cassou, 2015; Stephenson et al., 2006; Reintges et al., 2017; Ulbrich et al., 2008). 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)~~ according to Moore et al. (2013) . These authors 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 “pure”’ 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. ~~4~~5). The second point may be targeted by a combination of ~~correlation analysis~~ the strength of the responses ~~and the α_1 values~~ (correlations r) and the response values themselves (α_1): NAO responses in the CEUR domain of all data sets are most reliable in regions where 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~~ However, linearity 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 in these regions is only to a small fraction influenced by the NAO in this data set and period under consideration. In these cases, the NAO as expressed by the North Atlantic SLP gradient in this study is not the most important contributor and the noise, ε_Y in Eq. (4), is dominant.

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~~correlations. Thus, the future change of nSAT and PR per unit index change is most valid where ~~r is~~ correlations are high and where the ~~signals of α_1~~ NAO related responses emerge from internal variability, i.e. SNR > 1. Of course, α_1 and composite maps are not identical, as on the one hand the average index value that accompanies nSAT and PR anomalies is not the same (± 1 for α_1, but $+1.498$ and -2.103 for ~~REF~~ ERA-I composites, see Fig. ~~8~~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. (4)).~~.

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 ~~LBC~~ lateral boundary conditions is insufficient, but this is unlikely given the relatively small CRCM5 domain implying stronger ~~LBC~~ lateral boundary conditions 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. ~~4)~~ 5) and temporal correlation of GCM and RCM time series are generally high. Nevertheless, the influence of the lateral boundary conditions regarding SLP 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 ~~REF~~ response structures much finer than the CanESM2 and adds some robust high resolution geographical features which are clearly visible within the ensemble mean.

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. ~~5~~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. ~~7~~8) which shows a displacement in the GCM compared ~~to~~ with 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.

On the regional scale, the correlations in the CRCM5 are significantly stronger in several regions than in the CanESM2 (~~see Fig. 5~~Fig. 6–~~7~~8). These differences are not evened out by spatial aggregation~~(see Fig. 9).~~ . Thus, in the CRCM5-LE, more variance is explained by the NAO (i.e., by large-scale circulation) than in the CanESM2-LE. Explained variance is also higher in the single realizations of ERA-I and CRCM5/ERA-I than in the ensemble mean of GCM and RCM.

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 mean. 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.

When looking at spatially explicit ensemble sd maps (see Figs. ~~5–6and A4~~6–8 and A3), the ~~RCM LE~~ 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~~ with larger scales. However, the amplitude of the ~~IMS of r~~ inter-member spread of NAO–response correlations 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 ensemble values are shifted towards significantly higher r values in the RCM compared ~~to~~ with the GCM in both time frames, but not in the SE region.

~~When comparing present and future values, a vertical shift of the boxes in Fig. 9 indicates that r is~~

4.4 Climate Change

The results show 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 ~~, but the inter-quartile distance of the r distributions (box size) stays nearly the same for GCM and RCM. This shows that~~ climate simulation. Also the uncertainty range of the signals does not change significantly in the future horizon.

~~Temporally constant or only negligibly varying internal variability was already found for global mean temperature in Hawkins and Sutton (2~~ ~~assumed for global mean precipitation in Hawkins and Sutton (2011) .~~ With the here presented results, it can ~~also~~ be argued that the internal variability of more complex parameters (such as the NAO–response relationship quantified via Pearson correlation) shows no ~~siginificant~~ significant changes between historical and future periods.

When looking at the spatial distribution of α_1 ensemble 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 and the inter-member spread as stationary during these blocks such that the ~~IMS~~ inter-member spread 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.

4.5 ~~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.~~ The relative prevalence of negative index phases in the future period 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 be related with the emergence of negative index phases (Hanna et al., 2015; Woollings et al., 2010; Gillett and Fyfe, 2013; Cattiaux et al., 2013; Screen, 2017). 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~~ notes 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~~ with 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~~ with 1981–2010 in the entire Arctic and also the Barents/Kara Sea as ~~was~~ is verified with the CanESM2 variable “sea ice concentration” for this study (not shown).

An increasing frequency (relative to positive phases) of negative NAO events as noted in Fig. ~~8~~ 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 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 for nSAT mean, PR sum and nSAT ~~std~~sd. A coherent explication for this discrepancy might be that as correlations weaken, the Eurasian 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 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.

5 Conclusions

In this study, a RCM single-model initial condition large ensemble ~~was~~ is 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~~ is also included in the study. ~~Referring~~ With regard to the key questions raised in the introduction, it can be stated that:

(a) ~~Both large ensembles within the ClimEx project climate model chain~~ The ClimEx RCM-LE and its driving GCM-LE 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~~ individual realizations. The ensemble also shows ~~comparable climate statistics with the REF~~ climate statistics that are comparable with the reference time series and patterns. The clearly visible connection of the NAO with nSAT mean and PR sum follows well-known patterns ~~. The~~ and allows to derive robust information on the influence of the NAO on nSAT variability ~~, as expressed by the analyses on nSAT std, is also remarkable~~(nSAT sd).

(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~~ by 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~~ with 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~~ 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~~analyzed GCM-RCM combination, they allow drawing some general conclusions. The results strengthen the validity of ~~the climate module~~ this GCM-RCM combination for further applications, as important large-scale teleconnections only present in the GCM propagate properly to the fine-scale dynamics in the RCM. 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

~~Spatial patterns of change in nSAT 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 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.~~

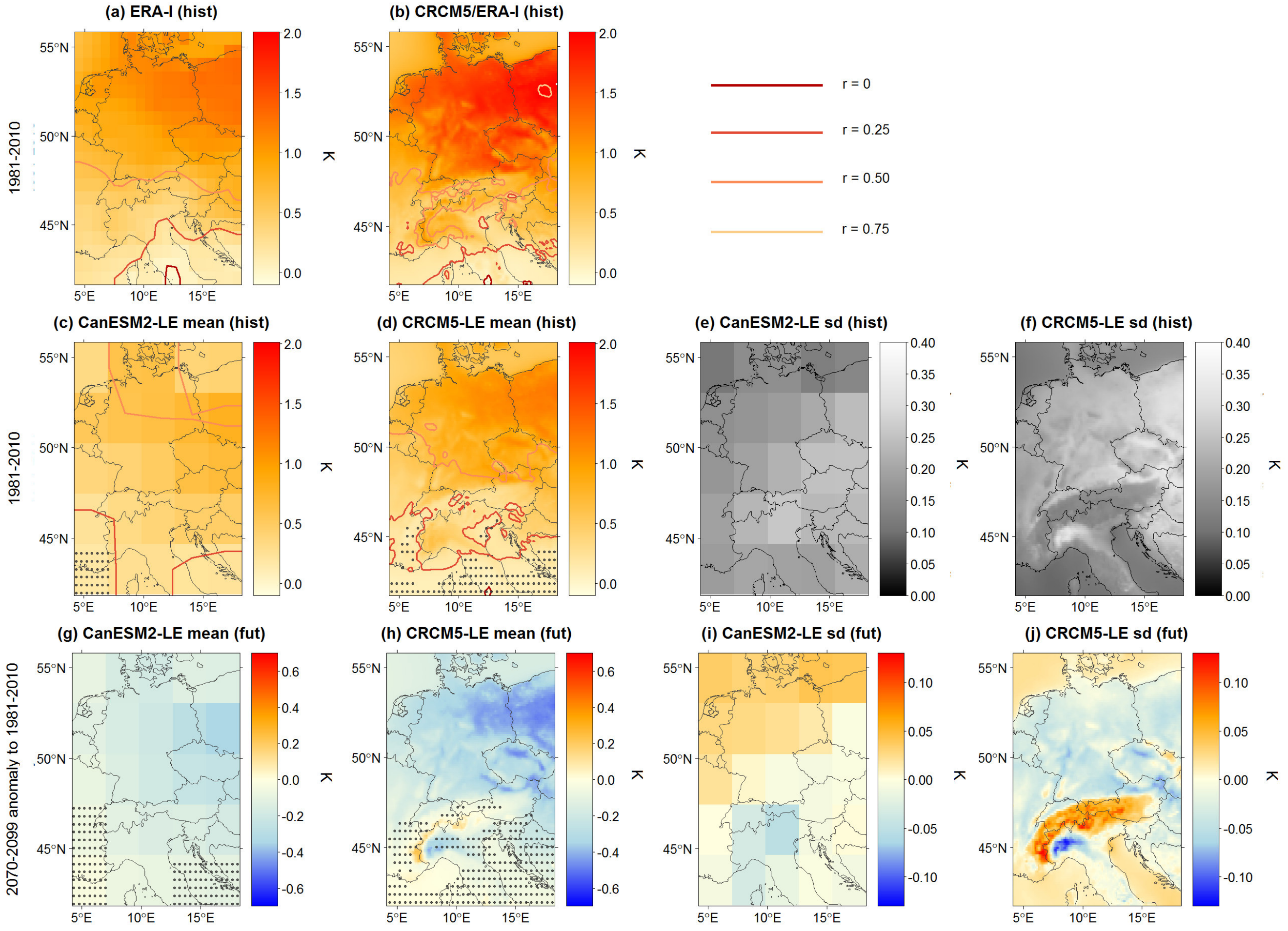


Figure 6. ~~Like Fig. 5, but for~~ Spatial patterns of change in nSAT ~~std~~ mean (α_1 in [K]) ~~. Note that~~ for a unit change in the ~~difference maps~~ NAO index for ERA-I, CRCM5/ERA-I, CanESM2-LE and CRCM5-LE in 1981–2010 ((a)–(f)) and the change in 2070–2099 with respect to 1981–2010 ((g)–(j)). Both 50-member ensembles are represented with ensemble mean ((c)–(d), (g)–(h)) and standard deviation (sd, (e)–(f), (i)–(j)) representing the inter-member spread. Reddish lines in the ensemble mean maps represent the Pearson correlation between nSAT mean and the NAO index at an increment of 0.25; red shadings see legend in upper right panel. Grey stippling in the 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. Stippling will be explained in more detail in Section 3.2.2.

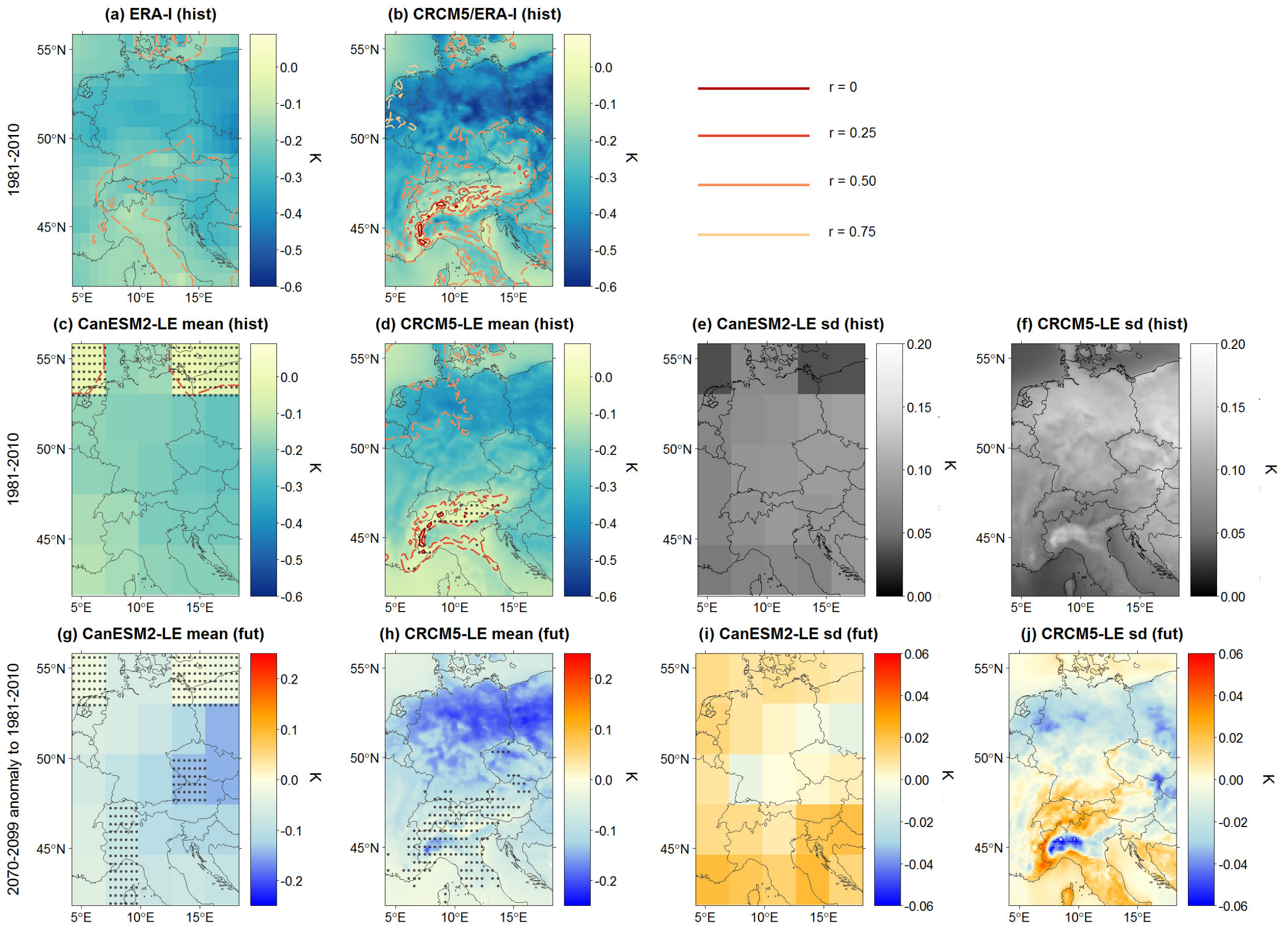


Figure 7. Like ~~Figs~~Fig. ~~5–~~6, but for ~~PR sum~~ nSAT sd (α_1 in [K]). Dashed lines of correlation coefficients indicate negative values. Note that the difference maps for CanESM2-LE and CRCM5-LE mean were calculated using absolute values~~and that the colour bar in the bottom row is flipped compared to Figs.5–6.~~

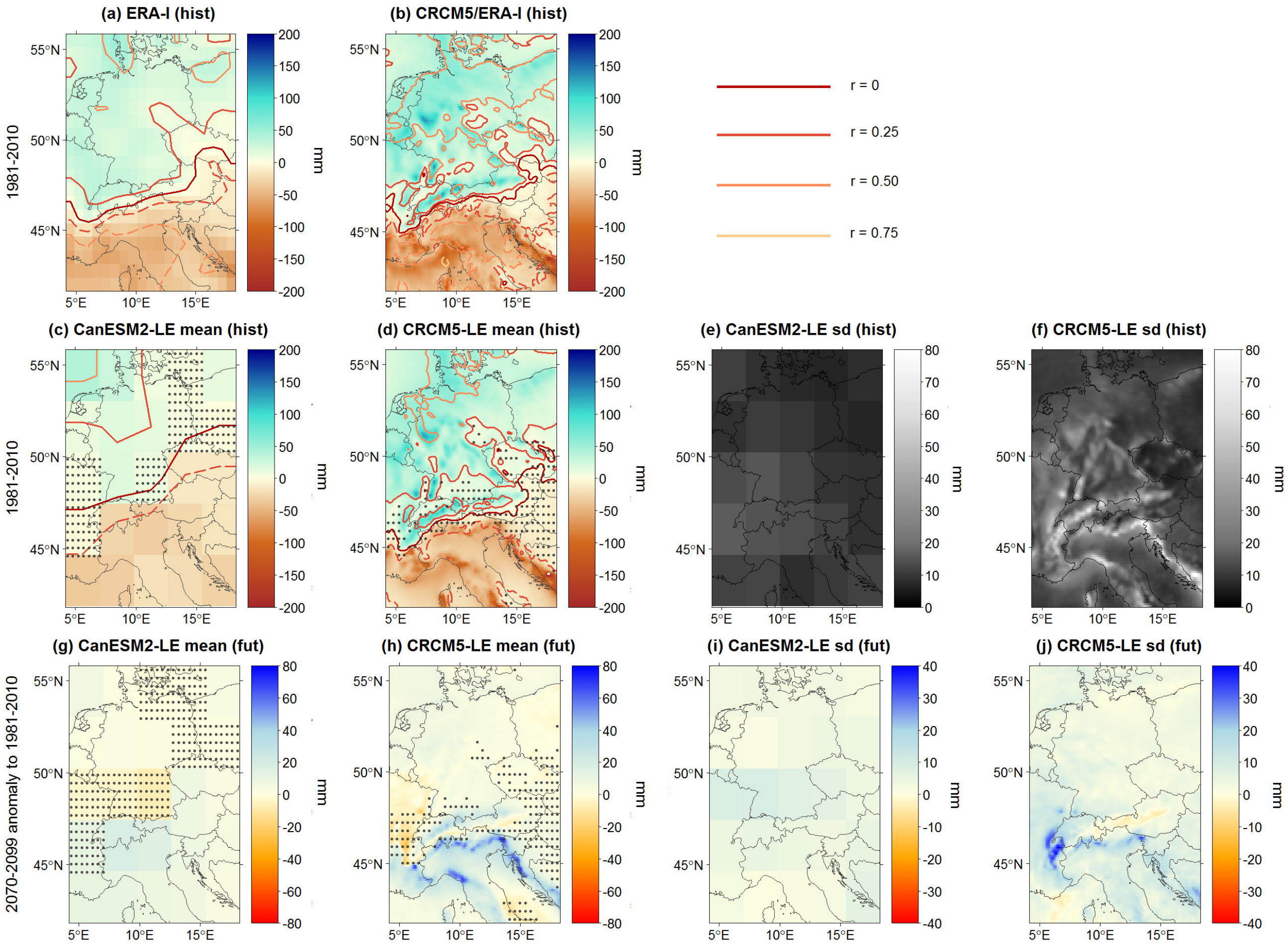


Figure 8. Like Figs. 6–7, but for PR sum (α_1 in [mm]). Dashed lines of correlation coefficients indicate negative values. Note that the difference maps for CanESM2-LE and CRCM5-LE mean were calculated using absolute α_1 values and that the colour bar in the bottom row is flipped compared with Figs. 6–7.

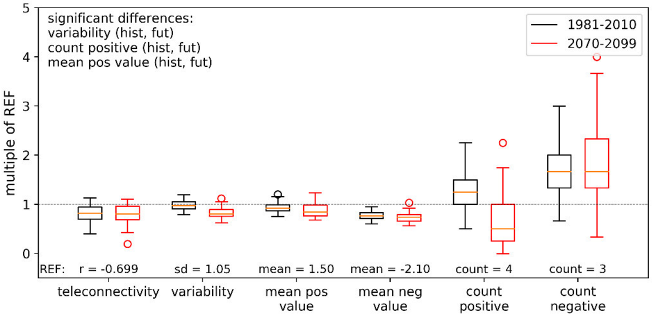


Figure 9. Several index statistics of all 50 CanESM2-LE members expressed as multiples of the respective ERA-I value (~~REF~~ ERA-I value set to 1.0): teleconnectivity (Pearson correlation between AH and IL time series), index variability (expressed as temporal standard deviation ~~in time~~ of index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

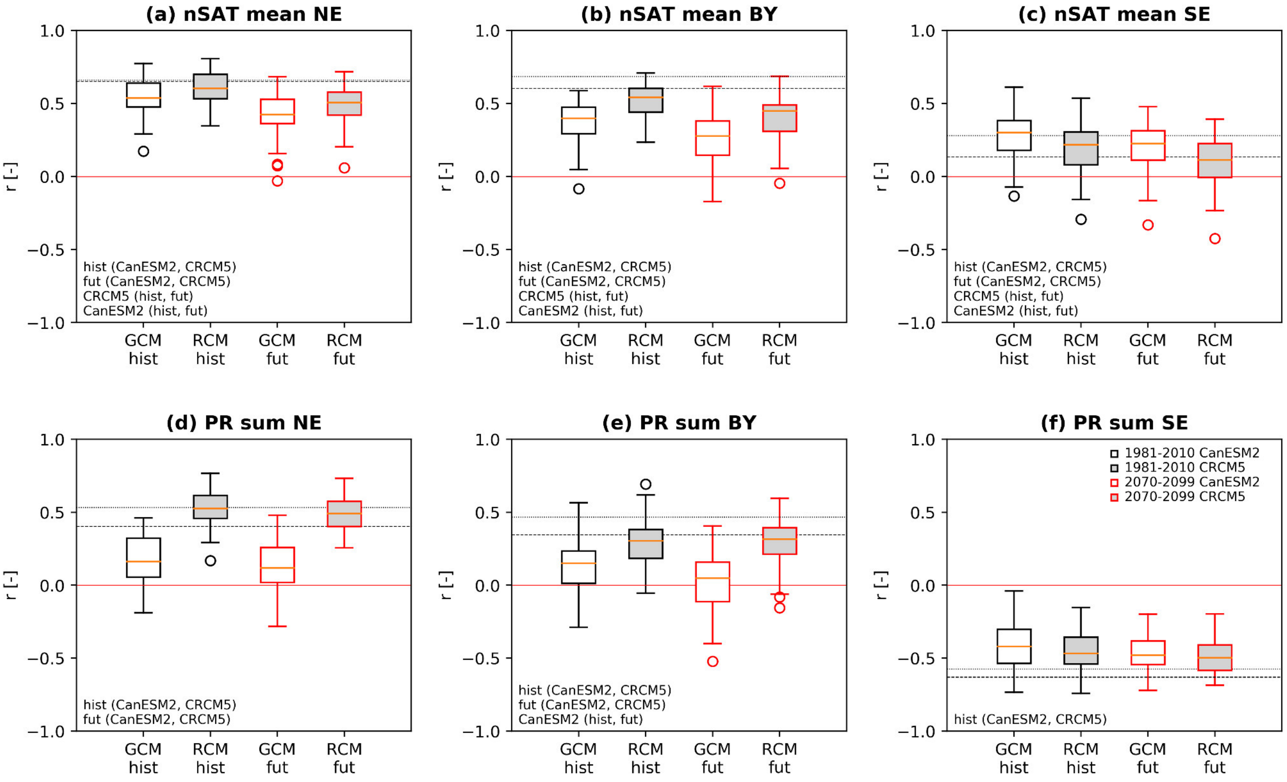


Figure 10. Boxplots of nSAT mean ((a)–(c)) and PR sum ((d)–(f)) showing Pearson correlation of 50 CanESM2-LE (white filling) and CRCM5-LE (grey filling) realizations for three regions (NE, BY, SE) in historical (black outlines) and future (red outlines) time horizons. Dashed (dotted) horizontal lines indicate the ERA-I (CRCM5/ERA-I) value; text denotes combinations of which the differences are significant 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~~ CanESM2-LE and ~~RCM LE~~CRCM5-LE. Solid orange line in boxplots: median. For regions NE, BY, SE see Fig. 2.

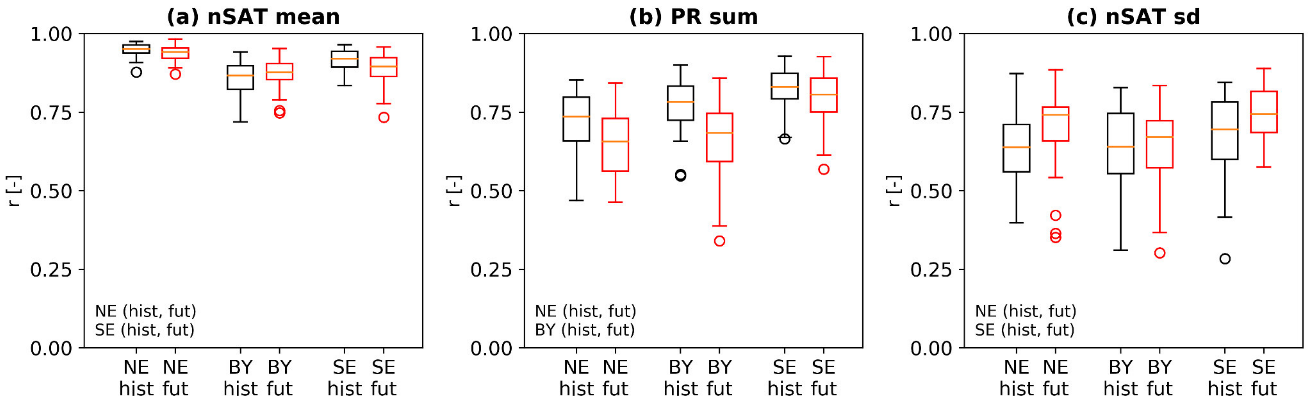


Figure 11. Temporal co-variability of ~~CanESM2~~ CanESM2-LE and ~~CRCM5~~ CRCM5-LE subset regions in all 50 members. Each boxplot represents 50 Pearson correlation coefficients between the time series of variables nSAT mean (a), PR sum (b) and nSAT ~~std~~ sd (c) in the subset regions of ~~CanESM2~~ CanESM2-LE members and the corresponding ~~CRCM5~~ CRCM5-LE members. Time periods used for correlations: 1981–2010 (hist, black), 2070–2099 (fut, red). For regions NE, BY, SE see Fig. ~~1~~2. Text denotes combinations of which the differences are significant at $p \leq 0.05$ using an unpaired Mann-Whitney/U-test. Solid orange line in boxplots: median.

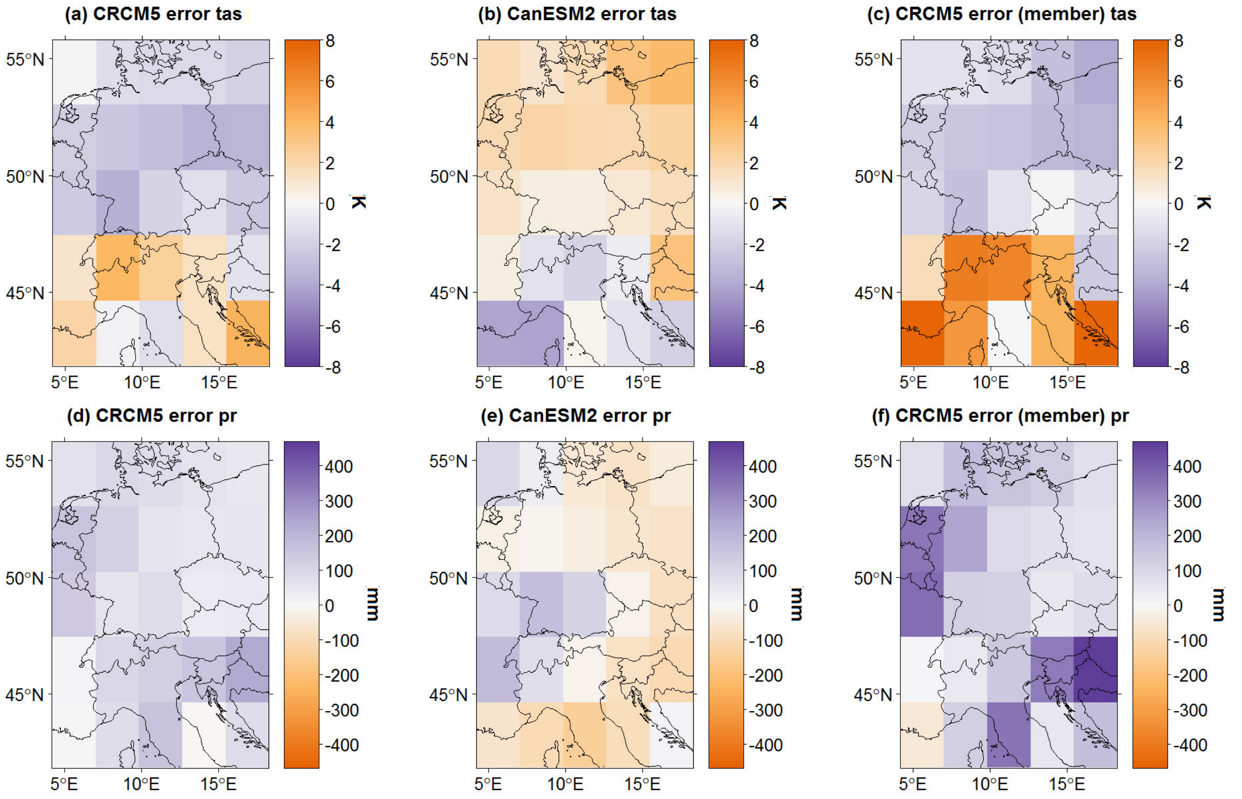


Figure A1. Model ~~errors~~ deviation for ~~30~~ the 1981–2010 winter mean nSAT mean ((a)–(c)) and ~~30~~ winter mean PR sum ((d)–(f)) in GCM resolution (2.8°). First column: error of CRCM5 under ~~“perfect”/~~ERA-I boundary conditions (difference between CRCM5/ERA-I and ERA-I). Second column: error of ~~GCM~~ CanESM2-LE towards ERA-I data (ensemble mean of differences between ~~GCM~~ CanESM2-LE members and ERA-I). Third column: CRCM5 error under ~~GCM~~ CanESM2-LE boundary conditions (ensemble mean of differences between CRCM5 members and corresponding CanESM2 members).

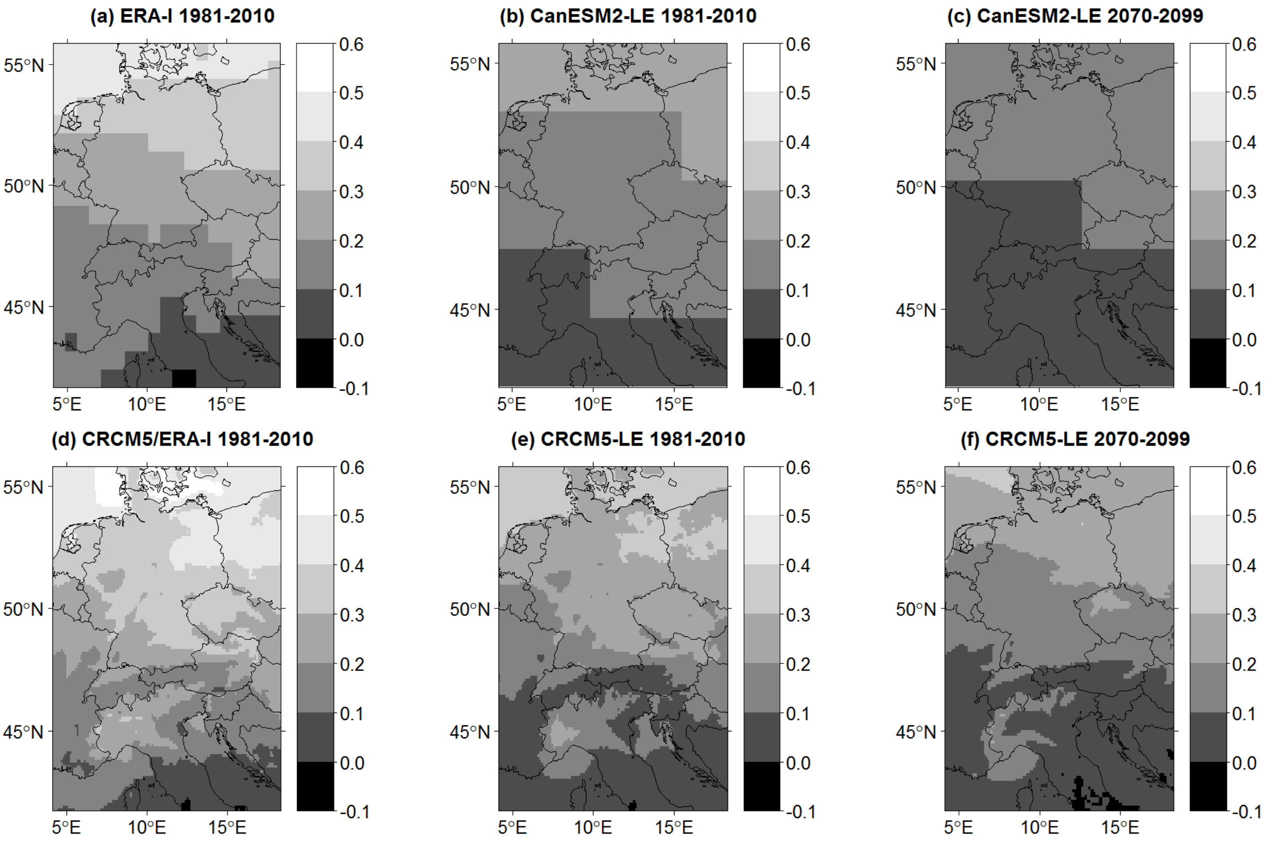


Figure A2. Ratio of nSAT α_1 and winter mean daily standard deviation of nSAT for ~~CRCM5/ERA-I~~ driving data ((a)–(c)) and ~~ERA-I~~ RCM data (~~b~~(d)~~under~~ –(f)) during historical ~~conditions and CanESM2-LE mean~~ ((~~e~~a)–(b), (d)–(e)) and ~~CRMC5-LE mean~~ future ((~~e~~c)~~–~~, (f)) ~~under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

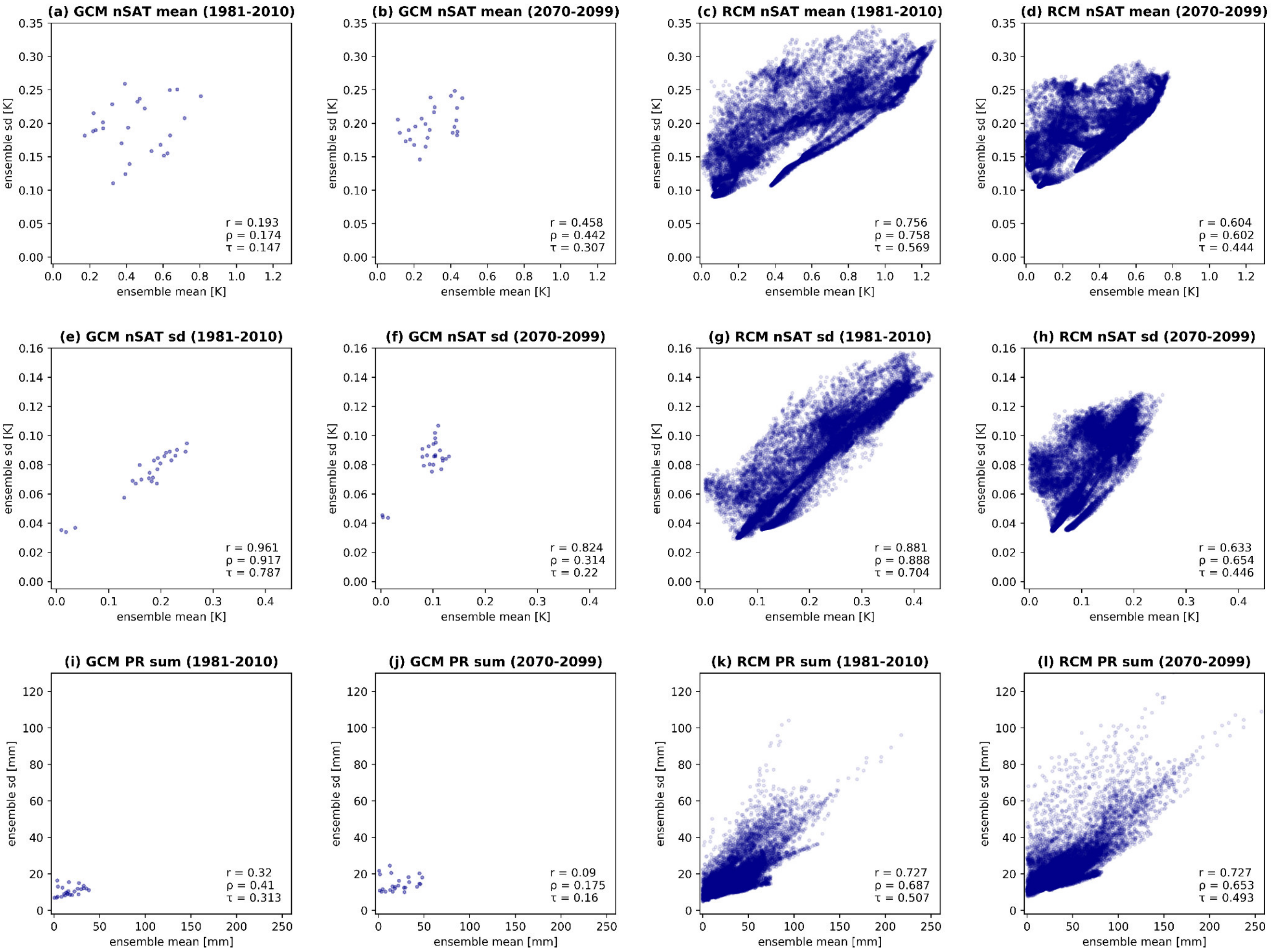
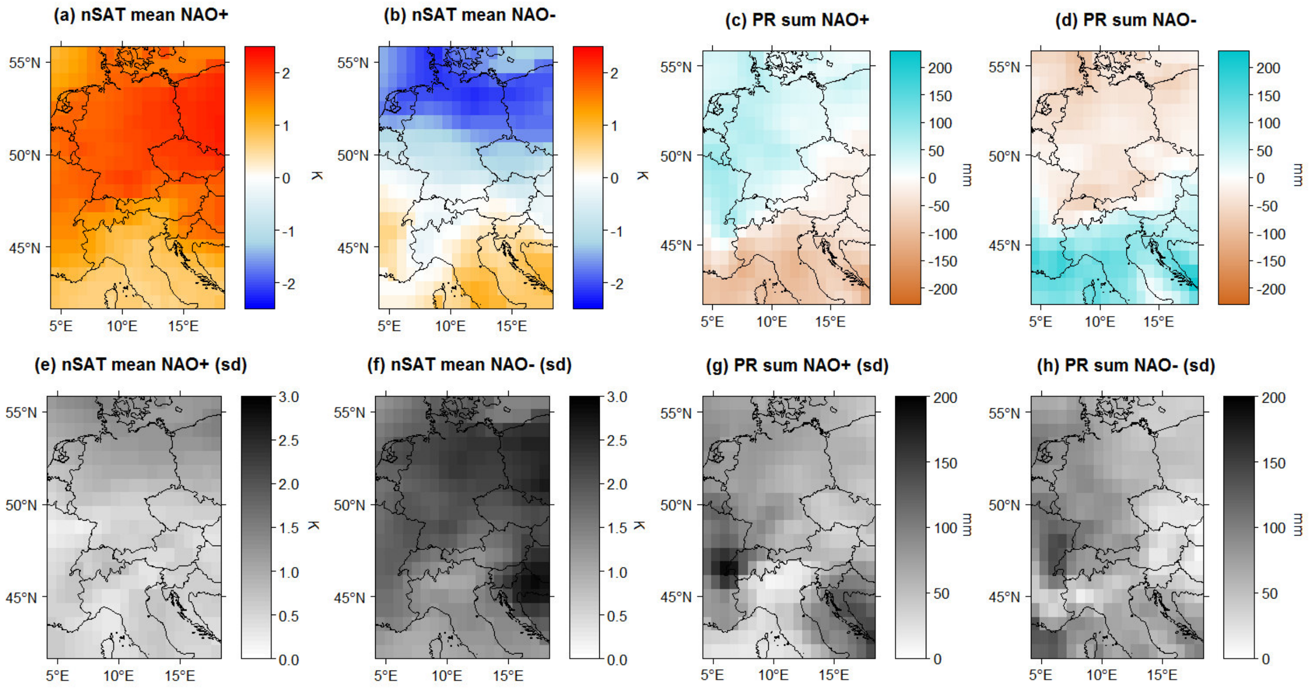


Figure A3. Relationship between LE mean and sd values for variables nSAT mean (a)–(d), nSAT ~~std~~ sd (e)–(h), PR sum (i)–(l) for hist and fut ~~time frames~~periods. ~~Upper~~ Lower right corner: r – Pearson correlation coefficient, ρ – Spearman rank correlation coefficient, τ – Kendall's Tau.



(a) nSAT mean NAO+

(b) nSAT mean NAO-

(c) PR sum NAO+

(d) PR sum NAO-

(e) nSAT mean NAO+ (sd)

(f) nSAT mean NAO- (sd)

(g) PR sum NAO+ (sd)

(h) PR sum NAO- (sd)

Figure A4. ERA-I anomalies from the long-term mean of nSAT mean in $[\mathrm{K}]$ and PR sum in $[\mathrm{mm}]$ in NAO positive (1989, 1990, 1994, 1995) and negative (1996, 2001, 2010) winters. Mean index value for positive (negative) NAO phases is $+1.498$ (-2.103).

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~~ which helped to considerably improve 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

Aalbers, E. E., Lenderink, G., van Meijgaard, E., and van den Hurk, B. J. J. M.: Local-scale changes in mean and heavy precipitation in Western Europe, climate change or internal variability?, Climate Dynamics, 50, 4745–4766, https://doi.org/10.1007/s00382-017-3901-9, https://doi.org/10.1007/s00382-017-3901-9, 2018.

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.

Frankcombe, L. M., England, M. H., Mann, M. E., and Steinman, B. A.: Separating Internal Variability from the Externally Forced Climate Response, Journal of Climate, 28, 8184–8202, https://doi.org/10.1175/JCLI-D-15-0069.1, https://doi.org/10.1175/JCLI-D-15-0069.1, 2015.

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.

Kay, J. E., Deser, C., Phillips, A., Mai, A., Hannay, C., Strand, G., Arblaster, J. M., Bates, S. C., Danabasoglu, G., Edwards, J., Holland, M., Kushner, P., Lamarque, J.-F., Lawrence, D., Lindsay, K., Middleton, A., Munoz, E., Neale, R., Oleson, K., Polvani, L., and Vertenstein, M.: The Community Earth System Model (CESM) Large Ensemble Project: A Community Resource for Studying Climate Change in the Presence of Internal Climate Variability, Bulletin of the American Meteorological Society, 96, 1333–1349, https://doi.org/10.1175/BAMS-D-13-00255.1, https://doi.org/10.1175/BAMS-D-13-00255.1, 2015.

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.

Maher, N., Matei, D., Milinski, S., and Marotzke, J.: ENSO Change in Climate Projections: Forced Response or Internal Variability?, Geophysical Research Letters, 45, 11,390–11,398, https://doi.org/10.1029/2018GL079764, 2018.

Maher, N., Milinski, S., Suarez-Gutierrez, L., Botzet, M., Dobrynin, M., Kornblueh, L., Kröger, J., Takano, Y., Ghosh, R., Hedemann, C., Li, C., Li, H., Manzini, E., Notz, D., Putrasahan, D., Boysen, L., Claussen, M., Ilyina, T., Olonscheck, D., Raddatz, T., Stevens, B., and Marotzke, J.: The Max Planck Institute Grand Ensemble: Enabling the Exploration of Climate System Variability, Journal of Advances in Modeling Earth Systems, 11, 2050–2069, https://doi.org/10.1029/2019MS001639, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019MS001639, 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.