Responses to Referee Comment No. 2

Dear Reviewer No. 2,

We would like to thank you very much for your constructive and valuable comments, particularly the very interesting suggestions for further analyses. We think that they will considerably improve the scientific quality of our manuscript. Below, we answer to all points raised in the Referee Comment (comments in grey, answers in black). Please consider also the responses to Reviewer No. 1 as there are some cross-overs regarding the comments.

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

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

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

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

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

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

The lack of (structural) clarity in the manuscript is also criticized by Reviewer No. 1, and we will work on the manuscript to address this issue.

Major Comment:

RCM reproduction of NAO teleconnection in driving data

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

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

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

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

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

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

RMS (i,j) = Ave_(m=12,1-3){Ave_n { SQRT[Ave_t {D_m(i,j,t,n)^2 }] / Var_Drive_m(i,j,n)}}

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

Var_Drive_m(i,j,n) = Ave_t {[SLP_Drive_m (i,j,t,n) - Ave_t {SLP_Drive_m (i,j,t,n) }]^2}.

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

will be calculated on the entire ClimEx domain. We expect the RMS to be quite low in the entire domain, and we will conduct a significance test on H0: RMS(i,j) >=1 (maybe also 0.5 or even lower).

Minor Comments:

l.2 "natural variability". Later it seems, "internal variability" (l. 16) is used to refer to the same phenomenon. It would be helpful to be consistent throughout.
This is true, thank you. We will use "internal variability" throughout the study.

l. 5-6. "its transfer from the driving model CanESM2 into the driven model CRCM5." Perhaps better wording might be "its representation in the driven model CRCM5 relative to the driving model CanESM2."
Thank you, we will adopt the wording suggestion.

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

l. 21 "is to apply slight differences in" -> "is to perturb"
ll.21-22 "with similar long-term climate statistics" This refers to a response rather than an experimental setup. I think it might be more correct to say "under identical external forcings"
l. 44 "its dynamics in a future climate" -> "its fidelity in a future climate"
Thanks, we will consider the phrasing suggestions for lines 21, 21-22, 44.

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

ll. 65-66 "finding robust NAO patterns which exceed the uncertainty due to internal variability in the ensemble." The phrase, "exceed the uncertainty due to internal variability" is confusing in this context. Perhaps say, "finding robust NAO patterns by significantly reducing sampling uncertainty associated with internal variability"
The suggested formulation is certainly clearer than the original one. We will adopt the suggestion.

l. 71 "Do GCM NAO impulses propagate correctly into the RCM realizations" perhaps better stated as, "Does the RCM correctly represent the NAO pattern present in the driving data" (ie my major comment)
We agree that key question (c) is better stated in this way as the suggested wording also encompasses the additional analyses regarding the large-scale SLP pattern. We will rephrase the sentence.

ll. 68-74. These are excellent focal points/topics for the paper. It would be very helpful if these were better referred back to in the analysis, discussion, and summary sections so the reader can more easily keep track of which of these you are addressing and what progress you have made on each.
Thanks! We will structure sections 3—5 accordingly. Please have also a look at Referee Comment 1 regarding the Introduction and Conclusions sections.

ll.101-103. two names are presented for each of three variables (eg msl/psl, t2m/tas, and tp/pr). I did not see a reason for this. If there is a reason it should be stated. If there isn't, then it would be clearer if just one name was presented for each and used throughout the paper.
Thanks for this note. Please have also a look at the responses to Referee Comment 1 where we address a similar issue. The two names refer to different model variable output names (e.g. msl, t2m, tp were derived from ERA-I, psl, tas, pr from CanESM2 and CRCM5). We will also change the analysis variable names like psl → SLP, tas mean/std → nSAT mean/std, pr sum → PR sum in the text.

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

Thanks for this hint. We will replace Table 2 with a map which will indicate all domains employed in the study. We think this is a better way to illustrate the position and extent of the domains than listing the boundary coordinates.

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

Yes, this is certainly true. In combination with the following suggestion we will change the colors and possibly also the diagram style (CDFs instead of histograms).

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

This is a very nice idea. We will consider an analysis following these steps. In order to better discriminate the different groups (and periods) we will also switch from histograms to CDFs. Names of the two groups may be SOIC – "same ocean initial conditions" (looking at members from the same family), and MOIC – "mixed ocean initial conditions" (looking at members from different ocean families) as in Leduc et al. 2019.

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

When correlating the ERA-I realization with the 50 CanESM2 members we were not so much looking for a control group. The idea was to evaluate whether ERA-I can be seen as another independent realization, but with comparable climate statistics. We will include a short explanation on the reasoning in the text.

l. 214 "positive, negative and indifferent index values" -> "positive, negative and neutral index values"

Thanks, we will change "indifferent NAO states/index values" to "neutral" in the text. Please have also a look at Referee Comment 1.

l. 223 "it backs the choice" -> "it supports the choice"

Thanks, we will change it.

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

We agree that there is a lot "back and forth" which is related to the fact that we tried to respect the strict separation of the results and discussion sections. In order to improve readability though and following some comments of Referee Comment 1, we will restructure the discussion with respect to the four key questions, integrate the results and remove too detailed repeats of results.

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

Thanks, we will change the formulation.

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

This is an interesting connection. We will look into the data to assess it. However, we have already established that low correlations (weak NAO relationships) do not occur in the same locations for all variables (e.g. for pr sum, they are mostly located near the Alps; but for tas mean they are found in the southern part of CEUR, where pr sum shows relatively strong correlations with the NAO).

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

We will also analyze the RMS results to assess the SLP representation between GCM and RCM.

Reference:

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