CC1#1: *Evin et al. present an impressive and extensive study involving the ANOVA method and data augmentation to estimate balanced mean change with an approach called 'QUALYPSO' and to distinguish between sources of uncertainties: RCPs, GCMs, RCMs and internal variability. This is a contribution to progress in terms of understanding downscaled model projections and the use of ensembles.*

Thank you for this positive comment.

CC1#2: I nevertheless have some suggestions concerning the background and introduction to this study. My impression is that there are many papers presenting RCM results that ignore related work done through empirical-statistical downscaling (ESD). I think that acknowledging such work in many cases would strengthen them. Here for instance, the statement "the largest MME of regional climate projections ever produced" is not quite correct since ESD efforts for a while have produced larger Multiscenarios Multimodel Ensembles (MME). For instance, a total of 254 downscaled simulations, each for 200 years, provided a basis for new ways to present large ensembles in Benestad et al. (2017; DOI: 10.1016/j.cliser.2017.06.013).

Thank you for this comment. We totally agree that ESD approaches have their merits and are a valid alternative to dynamic downscaling approaches provided by RCMs. This is indeed missing in the first version of the manuscript and will be added in introduction and in conclusion.

Concerning the introduction and the reference to *"the largest MME of regional climate projections ever produced"*, we will precise that we are here referring to "the largest MME projections based on regional climate models ever produced".

CC1#3: I would also argue that any effort to provide a robust estimate of total uncertainty in connection with downscaling should involve both RCMs as well as ESD, because these two approaches have different strengths and weaknesses independent of each other. They draw on different sources of information. For instance RCMs may be biased because of inconsistencies with the driving global climate model (GCM or ESM - here I use GCM referring to both). E.g. using different parameterisation schemes than the driving GCM, or there may be differences in outgoing longwave radiation (OLR) at the top of the atmosphere because the RCMs produce different rain/cloud climate to the driving GCM. Furthermore, both RCMs and ESD rely on the link between large and small scales being stationary, but in different ways: ESD in terms of the calibration of historical predictors and predictands; RCMs in terms of their parameterization schemes that provide a large-scale aggregation of unresolved small-scale processes.

Thank you for this comment. We are aware of the strengths and weaknesses of ESD and RCM approaches. In addition, we agree with you that complementary between dynamical and statistical approaches is a key to provide useful regional climate information. In our opinion, the systematic inclusion of projections based on ESD can be discussed. Indeed, bias correction techniques or data augmentation methods can also be relevant depending on the application. In the case of our study, we will underline the complementarity between statistical and dynamical approaches and a new sentence will be added in conclusion. A detailed discussion of the strengths and weaknesses of ESD and RCM approaches seems to be out of the scope of the study.

CC1#4: A comment on "the Arctic warming amplification and to the regional snow-albedo positive feedback" is that the Arctic amplification is even more pronounced at the high *latitudes, during winter, when there is no sunlight (during the polar nights). It's not so obvious that this is due to an albedo effect because it's dark (and there is also often a cloud-cover present).*

Concerning the Arctic warming amplification, we remind that the full sentence in the current manuscript is "*In winter and with the scenario RCP85, NEU and CEU are warming substantially more than the MED area likely due to the Arctic warming amplification and to the regional snow-albedo positive feedback*". This is a misunderstanding here, both processes (Arctic amplification and snow-albedo feedback) are mentioned as 2 possible processes able to explain the extra warming in the North and Central Europe (NEU, CEU) but they are not linked in our phrasing. In particular we confirm that the snow-albedo feedback can play a role over land in Central Europe (CEU) in Winter.

CC1#5: It's a bit surprising that the internal variability converges to zero at 2100 as it seems like in Figure 2 - column 5 when the total temperature change only is a few degrees C (especially for northern Europe). I suggest checking the calculations. Also see Deser et al. (2012; DOI: 10.1038/nclimate1562).

Internal variability of a modelling chain at a given time typically refers to the variance of all realizations that could be obtained with that chain at that time from multiple runs. In our work, it is estimated as the (temporal) variances of 30-years average deviations from the estimated climate change response of the chain (see Eq. S10 in the SM) and the internal variability estimate of the multimodel EURO-CORDEX ensemble is estimated as the multichain mean. It is thus assumed to be constant over time and does consequently not converge to zero.

For mean temperature changes in winter, for region CEU, internal variability is equal to 0.023 [°C^2] as a variance and 0.15 [°C] as a standard deviation. At the end of the century, the total uncertainty variance (including internal variability) is equal to 1.86 [°C^2] as a variance and 1.36 [°C] as a standard deviation. This large total uncertainty variance is explained by the significant differences obtained between climate change responses for the different RCPs (in dark green in Fig. 2, column 5) and for the different GCMs (in dark blue in Fig. 2, column 5). At the end of the century, the contribution of internal variability (as a **variance**) to the total uncertainty (again as a variance) is thus very small (around 1.2% of the total variance). This small contribution corresponds to what is presented in column 5 of Figures 2 and 4.

As mentioned in the manuscript, note that the estimate of internal variability would have been larger for 20-yrs, 10-yrs or 1-yr averages (see l. 301-307 of the current manuscript) but it would not have been larger than 2 times for 20-yrs averages and 3 times for 10-yrs averages because of the expected temporal correlation in 10-yrs averages (likely induced by internal variability), see the additional results in response to the comment RC1#6 below.

It is difficult to compare our results to those presented in Deser et al. (2012), the data and the statistical framework used for the estimation differing significantly. Estimates from Deser et al. 2012 are obtained from a 40-member ensemble of 57 years (2003-2060) obtained with one GCM for one single emission scenario (SRES A1B), and internal variability is estimated from the standard deviation of 8-year low-pass filtered data where the low-pass filter is a 5-point binomial filter. Internal variability can vary significantly from one model to the other and as mentioned above, internal variability also depends on the aggregation scale of the studied variable (8-yrs aggregation in Deser et al. 2012 compared to our 30-yrs aggregation). Even with these differences, the internal variability deviation estimate in Deser et al. 2012 is coherent with ours : it varies in Europe from below 0.3 (Southern, Central) to below 1.2°C (Scandinavia) in DJF and is below 0.9°C (below 0.3 in Scandinavia) in JJA (see Fig. 16 In Deser et al. 2012).

We will add some additional comments in the first paragraph where the internal variability is interpreted (current lines 189-203) and in Section 6 “Main contributions to the total variance” (lines 301-307).

CC1#6: GCM uncertainty over sea may be a result of incorrect sea-ice cover since the temperature shoots up in the air where it retreats, as mentioned in the discussion. This has been interpreted as a known shortcoming in the past, and it’s a bit surprising if the models with sea-ice in this region also are considered among the most trustable CMIP5 GCMs concerning the wintertime sea-ice cover. I suggest checking this.

GCM uncertainty over the sea can be related either to the SST change pattern or to the sea ice cover (SIC) pattern indeed. We agree that the presence or absence of sea ice can strongly modify the near-surface atmosphere temperature and therefore explain locally the wintertime GCM uncertainty. However SIC can play a key role without invoking incorrect sea-ice cover. Indeed even with “good” GCMs, the uncertainty in the sea ice cover response is expected to be strong. To carefully check this point, we would need to access the SST and SIC of both the GCMs and RCMs. To our knowledge, this information is not available for RCMs. This said, it is still possible that GCMs with implausible sea ice cover behaviour are part of the EURO-CORDEX driving GCMs as we are not aware that the driving GCMs have been selected specifically on this criterion. This is however planned for the selection of the CMIP6 GCMs.

CC1#7: Uncertainties should probably not exclude the possibility of tipping points. In this case, it could be a reversal of the thermohaline circulation.

We fully agree with the reviewer. GCM simulations with tipping points should not be excluded a priori if they can be considered as plausible futures. It is not certain however, that “reversal of the thermohaline circulation” can be considered as plausible during the 21st century. Nevertheless, QUALYPSO method relies on the existing GCMxRCMxRCP matrix. It means that it cannot be used to extrapolate towards GCMs that have not been downscaled by any RCM in the initial matrix. So to include GCMs with tipping points in the QUALYPSO uncertainty range, we need to have them chosen by EURO-CORDEX. This is a limitation of the study and we will add the following paragraph in the discussion:

“*QUALYPSO method, as most of the matrix filling methods, relies on the existing MME. In our case, it means that it can not be used to extrapolate towards GCMs that have not been downscaled by any RCM in the GCMxRCMxRCP matrix considered in this study. If we want to generalize the combined use of dynamical downscaling and statistical approach to balance uncertainty estimates, we therefore advise the dynamical downscaling community to explore as much as possible a large diversity of plausible GCMs.*”

CC1#8: One suggestion: when it comes to precipitation, two key parameters are also the wet-day mean precipitation and the wet-day frequency. They are useful because they provide more actionable information than just the seasonal totals (their product with the number of days is the total precipitation amount).

Thank you for this comment. We agree that seasonal totals of precipitation are a very restrictive summary of the information provided by the climate projections available at fine temporal scales (e.g. daily). The same comment applies to temperature, seasonal averages being of limited interest for the most critical effects of climate change (e.g. intensification of heatwaves). 10-20 climate indices are customarily studied (see, e.g., Dosio et al., 2016). These analyses are obviously of interest and we expect to carry out such additional studies with QUALYPSO for part of them. Note that, however, for indices related to high quantiles, or involving a threshold, a bias-correction step is likely unavoidable as these indices represent characteristics of the statistical distribution that are typically poorly represented in raw climate projections.

CC1#9: Another suggestion is that methods from ESD can be used to study the connections between features provided by the driving GCM and the response simulated by nested RCMs. For instance, ESD calibrated with one GCM-RCM pair may be applied to a different GCM to compare with its RCM. This is a bit like 'hybrid downscaling'.

Thank you for this suggestion. We are well aware of the possibility to apply ESD-like techniques to develop hybrid downscaling. We however consider that this is out of the scope of the current study. The interest of SDMs and hybrid techniques to complete GCMxRCMxRCP matrices will be acknowledged in the future manuscript version, in conclusion ("*In addition to the dynamical downscaling approach provided by regional climate models, several alternative methods are available in the literature in order to improve the uncertainty assessment of future climate change. Empirical-statistical downscaling methods (Gutiérrez et al., 2019) have been proved to be a valuable approach to produce large ensembles representative of plausible futures (Benestad et al., 2017), statistical and dynamical downscaling approaches being seen as complementary within the EURO-CORDEX community (Jacob et al., 2020)*".

CC1#10: Finally, ESD can be regarded as a way to test the uncertainties connected with GCMs and decadal variability, and results by Mezghani et al (2019; DOI: 10.1175/JAMC-D-18-0179.1) may seem to suggest that internal variability plays a bigger role on a regional scale than the GCM (is didn't use 30-year smoothing, however). These results also highlight the limitation posed by 'the law of small numbers'. One nice aspect of ESD is that it can incorporate a quality evaluation of GCMs.

Thank you for this comment that again highlights the potential interest of ESD and will be acknowledged in the future manuscript version (see previous comment). Concerning the quality evaluation of GCMs, it is obviously another critical and key issue in climate impact studies. This issue will also be recalled in the manuscript. Indeed, the problem of the 'law of small numbers' (the fact that few RCMs and GCMs runs are available in some MMEs) is only partially resolved by QUALYPSO by balancing the estimates. However, obviously, it cannot take into account information about GCMs or RCMs that are missing from the MMEs, and the constitution of the MMEs is thus a critical step, independently of the uncertainty assessment approach (e.g. QUALYPSO), see our response to the comment CC1#7 above. Concerning the internal variability, we recall here, as it was already stated in the manuscript, that internal variability strongly depends on the spatial (pixel scale, averages over SREX boxes, countries, etc.) and temporal (interannual variability, averages over 10-yr or 30-yr periods) scales, as well as the variable analysed (temperature, precipitation, etc.), see our response to the comment RC1#5.

References

Benestad, R., K. Parding, A. Dobler, and A. Mezghani. 2017. “A Strategy to Effectively Make Use of Large Volumes of Climate Data for Climate Change Adaptation.” Climate Services 6 (April): 48–54. https://doi.org/10.1016/j.cliser.2017.06.013.

Deser, C., A. Phillips, V. Bourdette, et H. Teng. « Uncertainty in Climate Change Projections: The Role of Internal Variability ». Climate Dynamics 38, nº 3-4 (1 février 2012): 527-46. https://doi.org/10.1007/s00382-010-0977-x.

Dosio, A.. 2016. “Projections of Climate Change Indices of Temperature and Precipitation from an Ensemble of Bias-Adjusted High-Resolution EURO-CORDEX Regional Climate Models.” *Journal of Geophysical Research: Atmospheres* 121 (10): 5488–5511. https://doi.org/10.1002/2015JD024411.

Gutiérrez, J. M., D. Maraun, M. Widmann, R. Huth, E. Hertig, R. Benestad, O. Roessler, et al. 2019. “An Intercomparison of a Large Ensemble of Statistical Downscaling Methods over Europe: Results from the VALUE Perfect Predictor Cross-Validation Experiment.” *International Journal of Climatology* 39 (9): 3750–85. https://doi.org/10.1002/joc.5462.

Jacob, D., C. Teichmann, S. Sobolowski, E. Katragkou, I. Anders, M. Belda, R. Benestad, et al. 2020. “Regional Climate Downscaling over Europe : Perspectives from the EURO-CORDEX Community.” Regional Environmental Change 20 (2). http://urn.kb.se/resolve?urn=urn:nbn:se:smhi:diva-5677.