Interactive comment on "Development and prospects of the regional MiKlip decadal prediction system over Europe: Predictive skill, added value of regionalization and ensemble size dependency" by Mark Reyers et al.

Reviewer 1

General comments
This is a very good study that focuses on the potential merits of regional downscaling decadal climate predictions over Europe. Specifically, the MiKlip prediction system studied uses the low resolution MPI global decadal hindcast ensemble at T63 resolution and dynamically down scales these hindcasts over Europe using the COSMO-CLM model at 0.22° horizontal resolution. Two 10 member ensemble regional hindcasts of 5 start dates are examined and verified against observational analyses of surface temperature, precipitation and low level wind using three different skill metrics, MSESS, CRPSS and ACC. The authors examine these metrics to answer the following questions: is there potential for skillful regional predictions in Europe? Does regional downscaling provide added value? and How does the skill of these predictions depend on ensemble size? The first two questions are answered affirmatively and for the last question ensemble size stabilizes the skill metrics MSESS and CRPSS at ten members but ACC skill depends on ensemble size beyond ten members. The manuscript meets all the criteria for publication and needs only minor changes.

Specific comments:
The manuscript could be improved in two ways that would increase the significance of the work. First, although there are significantly large regions in Europe where the skill of the initialized hindcasts is positive, there is also a large region in central Europe where the skill is negative. This is particularly true of the MSESS of temperature. Since the reference is forecast is an uninitialized ensemble of 20th century simulations this raises the question as to the reason for this negative skill. The answer or some speculation to how it arises should be included in the article. In a similar vein, the authors do not include in their discussion any metrics that use the observed climatological distribution as the reference forecast, so that skill is measured solely using comparison with observations.

Answer: We thank the Reviewer for these helpful comments, which will certainly help to improve our manuscript. There are several potential reasons for the negative skill for temperature, including the detrending of the time series as queried by Reviewer 2. Following the suggestions of Reviewer 2 we will redo most of the analysis without detrending, which likely will result in new skillscore plots. The comparison of the new plots with those shown in the manuscript will probably help to better understand the reason for the negative skills, and we will discuss this in detail in the revised version when the new plots are available. Further, we followed the suggestion and performed an additional analysis including the climatology as reference. See for instance the upper row plots in Fig. A1-A3 attached to the answers to Reviewer 2 for ACC, which show the correlation w.r.t. the observations. The other skill scores (MSESS and CRPSS) have also been calculated for both references, namely the historical ensemble and the climatological distribution. We would include this information in the revised paper.

In Fig A1 you can see, that the skill in Central Europe is high when using the climatology as reference forecast (Fig. A1 upper left). But, in this region the historical ensemble shows an even higher skill (Fig.A1 lower left). Part of the reduced skill there can be attributed to the low sample size. Fig. A1 lower right indicates less negative to slightly positive skill scores compared to the historical ensemble when a larger sample size with annual starting values are used. We intend to include this additional analysis in the paper (see also answers to Reviewer 2).
Technical corrections
Pg 2 Yaeger et al should be Yeager et al
A: This will be corrected

Pg 8 stronger scattered should be more strongly scattered
A: We will change it accordingly