Earth System
Dynamics

Discussions

Open Access

EGU

*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.

Anonymous Referee #2

The paper gives a preliminary assessment of regional decadal prediction skill over Europe based on a high-resolution regional model forced with boundary conditions obtained from the low-resolution, global MiKlip prediction system. I deem the analysis preliminary because the "development and prospects" of the downscaling system are being assessed at a rather early stage when only 5 hindcast start dates have been completed using the regional model. This is a serious shortcoming that calls into question the reliability of skill scores (computed from 5 data pairs) that are used throughout to make statements about the benefits of downscaling for various fields in various Eu-

ropean regions. Two-tier decadal prediction involving regional downscaling is certainly a topic of high interest, but this manuscript has the feel of an internal technical note that documents some preliminary and very mixed results that are still clouded in uncertainty given the limited temporal sampling. Unless it can be shown (perhaps using the MPI baseline systems) that 5 start dates are sufficient to get an accurate estimate for the skill scores and fields of interest, then what is the point of all this? I suspect that 5 start dates is not sufficient, and that the skill scores reported here are very "noisy" as a result. This may contribute to the mixed results and lack of strong take-away messages from this paper. It may be better to wait until more downscaled start dates have been completed before resubmission of this analysis.

Another main concern is the use of detrending, which probably exacerbates the sampling issues (how well-defined is a trend computed from 5 data points?). There is no real need to detrend since you have an uninitialized ensemble that allows you to determine the skill improvement relative to the externally-forced signal (yes, pure ACC will be higher, but you can show delta(ACC), i.e. the change in ACC relative to the uninitialized ensemble).

The quality of the writing is decent, but not high, and there are numerous instances of poor English construction (some noted below). A thorough proofreading is in order if this is to be resubmitted.

Specific Comments and Questions:

P2,L8: Here and throughout: "Yaeger" should be "Yeager".

P2,L11: It's not clear what the point is of the "while few" construction. Are you contrasting the large number of studies focusing on global metrics with the relatively few studies focusing on storm tracks, etc? Please rewrite.

P2,L13: What is this an example of? Why cite Sutton and Hodson (2005) in a paragraph focused on initialized decadal prediction?

P3,L7-18: The motivation for the present work needs to be clarified, particularly since it is not at all clear how the present study differs from the closely related recent MiKlip studies that have just been cited (Kadow et al. 2016; Mieruch et al. 2014; Haas et al. 2016; Moemken et al. 2016).

P3,L11: This question is poorly phrased. Do you mean "depend on" the trend or "derive from" the trend?

P3,L15: This is a repetitive rephrasing of the questions just covered.

P3,L29: I don't understand how ocean temperature and salinity can be nudged towards NCEP/NOAA reanalysis, since the latter is an atmospheric reanalysis.

P4,L11: Not clear what is meant by "Analog to the global data"?

P4,L14: Replace "are" with "is".

P4,L16-19: You already introduced the ERA-driven CCLM simulation in the first line of this paragraph, so consolidate your sentences into one brief description.

P4,L21: I don't understand what you mean by "uninitialized model simulations started from historical CMIP5 runs". Do you mean downscaled simulations that can be considered "uninitialized" counterparts to CCLM_b0 and CCLM_b1? Do you mean "pre-industrial CMIP5 runs"?

P4,L32: Replace "the natural variability" with "natural variability". Why use linear detrending to isolate natural variability when you have just introduced an uninitialised ensemble that can be used to quantify the skill associated with external forcing?

P5,L14: I think you mean "post-processed time series".

P5,L24: What is the basis for claiming that "skill should originate mainly from the initialization" as opposed to the external forcing? This has not been shown and shouldn't be assumed.

P5,L25-: What are F(y) and Fo(y)? Please explain the CRPS equation. What exactly is CDF and how is it computed?

P6,L23-P7,L2: This is repetitive.

Fig 2: Suggest using a nonlinear scale for MSESS, such as -9 to 0.9 as in Shaffrey et al. (2016, doi:10.1007/s00382-016-3075-x), because this metric is not symmetric about 0 in terms of relative improvements in MSE. Please clarify that these are for annual mean (ie, not seasonal mean) anomalies.

P7,L7: I presume the detrending has been performed similarly for observations and for the uninitialized historical runs? This isn't explicitly mentioned.

Table 1: What is the meaning of "The uninitialized historical ensemble has been used as reference dataset", given that this is a table of ACC scores? Am I correct that this table displays correlations computed from 5 data points (corresponding to the 5 start years)? Clearly the externally-forced trend is important and so this table should include ACC scores for the uninitialized historical runs for comparison.

P7,L11: What is the meaning of "increases"âĂŤrelative to uninitialized or relative to detrended?

P7,L34: I would say Figure 2 shows more than a "slight shift".

P8,L2: Here and elsewhere delete "exemplary" as it is not being used properly.

P8,L4: It's curious that Fig 2e agrees so well with Fig 3a, but Fig 2f is so different from Fig 3b. Can you offer any explanation? In my mind, it calls into question the significance of skill scores computed from 5 data points.

P8,L25-29: This discussion begs the question of why you are doing any detrending at all (see comment above)? The purpose of the uninitialized ensemble is precisely to allow you to discriminate between greenhouse-gas induced variability (including trends) and natural variability (including AMO-related trends). Detrending is confusing matter-

sâĂŤjust compared initialized to uninitialized skill.

P9,L8: Change "whereas" to "and".

P9,L9-11: This incomprehensible sentence needs a rewrite.

P11,L5: I don't understand this sentence.

P11,L24-28: This is because you are doing bootstrapping without replacement; if you allow replacement, then the spread does not necessarily diminish with ensemble size.