Climate Prediction Workshop, Bergen, June, 2019

The skill of dynamical decadal forecasts with focus on the North Atlantic

Bo Christiansen and Shuting Yang

Danish Meteorological Institute, Copenhagen, Denmark

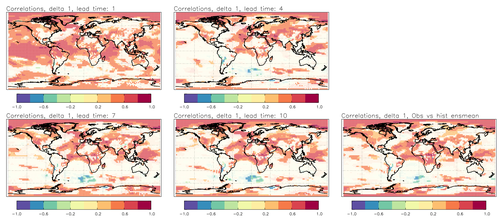
Introduction

We analyze the skill of decadal predictions based on ensemble experiments with dynamical climate models. The models are initialized every year in November for the period 1960-2005. We consider experiments with the Community Earth System Model (CESM) from the Large Ensemble Project and 3 CMIP5 models. The CESM ensemble has 40 members and the CMIP5 ensembles 10 members. For the considered models un-initialized (historical) ensembles with the same forcings exist. The advantage of initialization is analysed by comparing these two sets of experiments.

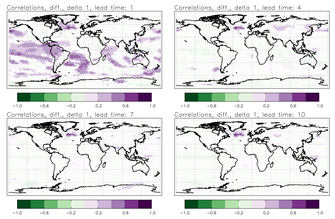
Here we focus on the near surface temperature (TAS) and use correlations between models and observations to estimate the skill. However, similar results are obtained using the root-mean square error.

The variability in both un-initialized and initialized experiments contains a considerable forced variability connected to green-house gasses and volcanic eruptions. This gives rise to predictability in both set of experiments. As the forced variability is difficult to remove entirely, we here compare the un-initialized and initialized experiments without any kind of detrending. Such a comparison requires careful considerations of the statistical significance in skills and, in particular, differences in skills. Here this is done by Monte-Carlo methods based on phase-scrambling the Fourier transform. When dealing with correlations between observations and models we randomize the time-series themselves. When dealing with the differences in correlations we randomize the model residuals obtained after removing the part of the variability linearly congruent with observations.

Correlations between ensemble mean and observations

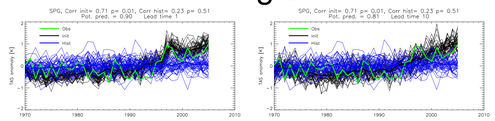


Strong correlations are found for lead-time 1 year. For longer lead-time good skill remain in many regions due to the forced variability. For longest lead-times skills are very close to the skill of the historical experiment. Large dots are where the correlations are significant to the 95 % level. LENS.

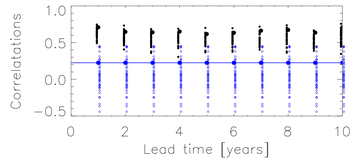


Difference between the correlations in the initialized and historical experiment. For the longest lead-times only little improvement by initializations is found except for the North Atlantic sub-polar gyre region. LENS.

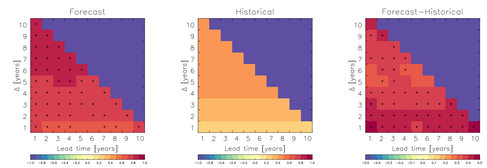
A closer look at the sub-polar gyre region



Time-series for averages over 40-15° W, 50-60° N. Black: forecasts for lead-times 1 and 10 years, Blue: historical, Green: observations. Thick curves: ensemble mean. Thin curves: ensemble members. The historical experiment is flat while the initialized forecast and observations are dominated by a jump around 1996. LENS.

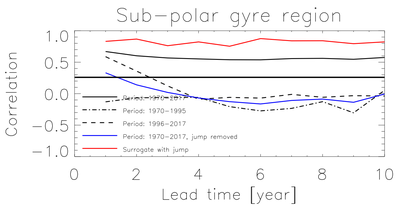


Correlations between model and observations as function of lead-time. Black: forecast, Blue: historical. Large dots: ensemble mean. Small dots: ensemble members. The ensemble members of the historical experiment have weak correlations between -0.5 and 0.5. For the forecast the correlations of ensemble members are around 0.5-0.8, depending only little on lead-time. For both historical experiment and forecasts the ensemble means have higher correlations than most of the individual ensemble members. See Christiansen 2018 (doi: 10.1175/JCLI-D-17-0197.1) and 2019 (doi: 10.1175/MWR-D-18-0211.1). LENS.



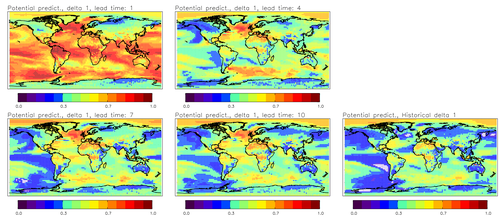
Correlation between ensemble mean and observations as function of lead-time and Δ (the number of years over which the predictands are averaged). There is a small increase in skill with increasing Δ for both forecasts and historical experiment. Dots indicate statistical significance on the 95 % level. No significant skills for the historical experiment. But for the forecasts and the difference between forecasts and historical experiment everything is significant. LENS.

Sub-polar gyre region: Deconstructing the skill



For the NA sub-polar gyre region large skill (correlation between ensemble mean and observations) is found for all lead-time (full black curve) for the full period. However, for the periods before and after 1996 weak or no skill is found (dashed black curves). Removing the jump in 1996 from observations by subtracting the means before and after results in weak skills for lead-times larger than 1 year (blue curve). Keeping only the jump in 1996 in the observations gives large skill for all lead-times (red curve).

Potential predictability



The potential predictability for the the initialized experiments (lead-times 1, 4, 7, and 10 years) and for the non-initialized experiment. The potential variability measures the ensemble mean spread relative to the total spread of the ensemble members. Note the difference between initialized and un-initialized experiments in the NA sub-polar gyre region. LENS.

Models and methods

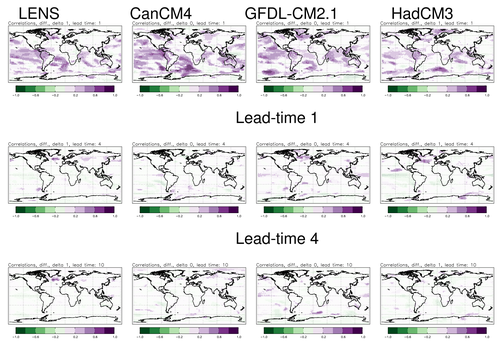
The models are initialized every year and have an ensemble of at least 10 members for both the decadal forecasts and the historical experiments.

Model	# Dec. Pred. members	# Historical members	Comments
LENS	40	42	The Community Earth System Model (CESM) Large Ensemble Project.
CanCM4	10	10	CMIP5.
GFDL-CM2p1	10	10	CMIP5.
HadCM3	10	10	CMIP5.

Observations: NCEP (2.5°x2.5°) or 20th Century Reanalysis V2 (NOAA) (2°x2°).

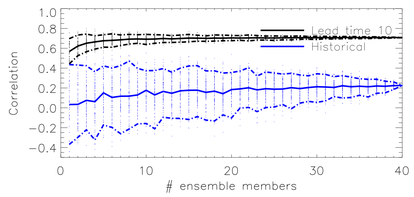
- We focus on near-surface temperature.
- No detrending. Compare historical experiments and forecasts to isolate influence of initializations.
- Simple bias-correction performed for each lead-time. Note that correlations do not depend on bias-correction.
- Forecast for 36 years: 1970-2005. Lead-times 1-10 years. Initializations done in November, a lead-time of 1 year means predicting the average over the year beginning the following January.
- Models interpolated to grid of observations.
- When comparing ensemble means: Use the same number of ensemble members.
- Statistical significance by Monte-Carlo methods which carefully take serial correlations into account.

Comparing models



Difference between the correlations in the initialized and historical experiments for all 4 models. Correlations are calculated between ensemble means and observations. Models agree on small effect of initializations except for the NA sub-polar gyre region.

Effect of ensemble size



The correlations between observations and ensemble mean for ensembles of different size. From the 40 member LENS ensemble sub-sets of ensembles of different sizes are randomly selected. The correlations between the ensemble mean of these sub-sets and the observations are calculated. The figure shows the mean (full curve) and 5 and 95 % quantiles (dashed curves) of these correlations. Black is initialized experiment for lead-time 10 and blue is historical experiment. The average skill improves with increasing ensemble size. However, large uncertainties in the skill of the ensemble mean are found, in particular for the historical experiment. Based on the spatial mean over NA sub-polar gyre region.

Conclusions

- Models agree that for lead-times between 4 and 10 years little effect of initialization is found except in the NA sub-polar gyre region. This (well-known) result is found across all the models and is robust to temporal and spatial smoothing.
- In the NA sub-polar gyre region the ensemble means of the forecasts explain 30-40 % more of the observed variance than the ensemble mean of the historical non-initialized experiments.
- However, the skill in the NA sub-polar gyre region seems to a large degree to be related to the shift towards warmer temperatures around 1996. Weak or no skill is found when only the sub-periods before and after 1996 are considered.
- The ensemble mean gives better skills than most individual ensemble members. The average skill improves with increasing ensemble size but large uncertainties are found. For some situations with large skill 10 ensemble members provides a good estimate. However, for situations with smaller skill the improvement with ensemble size is much slower and is not saturated for 10 ensemble members. Important to compare ensembles of same size.
- The potential predictability confirms the difference between the historical and the initialized experiments in the NA sub-polar gyre region.

The work was supported by the Horizon2020 project EUCP and BlueAction and by the NordForsk project ARCPATH. The Community Earth System Model (CESM) Large Ensemble Project and Decadal Prediction Large Ensemble Project were downloaded from `/net/isilon/ifs/arch/home/boc/LENS/` and `/net/isilon/ifs/arch/home/boc/DPLENS/`.