Useful decadal climate prediction at regional scales? A look at the ENSEMBLES stream 2 decadal hindcasts

D A MacLeod, C Caminade and A P Morse

School of Environmental Sciences, University of Liverpool, UK

E-mail: D.Macleod@liv.ac.uk

Received 18 June 2012
Accepted for publication 26 September 2012
Published 18 October 2012
Online at stacks.iop.org/ERL/7/044012

Abstract

Decadal climate prediction is a branch of climate modelling with the theoretical potential to anticipate climate impacts years in advance. Here we present analysis of the ENSEMBLES decadal simulations, the first multi-model decadal hindcasts, focusing on the skill in prediction of temperature and precipitation—important for impact prediction. Whilst previous work on this dataset has focused on the skill in multi-year averages, we focus here on the skill in prediction at smaller timescales. Considering annual and seasonal averages, we look at correlations, potential predictability and multi-year trend correlations.

The results suggest that the prediction skill for temperature comes from the long-term trend, and that precipitation predictions are not skilful. The potential predictability of the models is higher for annual than for seasonal means and is largest over the tropics, though it is low everywhere else and is much lower for precipitation than for temperature. The globally averaged temperature trend correlation is significant at the 99% level for all models and is higher for annual than for seasonal averages; however, for smaller spatial regions the skill is lower. For precipitation trends, the correlations are not skilful on either annual or seasonal scales.

Whilst climate models run in decadal prediction mode may be useful by other means, the hindcasts studied here have limited predictive power on the scales at which climate impacts and the results presented suggest that they do not yet have sufficient skill to drive impact models on decadal timescales.

Keywords: climate adaptation, decadal prediction, climate modelling

Online supplementary data available from stacks.iop.org/ERL/7/044012/mmedia

1. Introduction

The ability to forecast the evolution of climate on decadal timescales would prove a useful tool to aid climate change adaptation. Information on this timescale is important for adaptation as it is a key planning horizon for governmental and non-governmental organizations, businesses, and other societal entities (Cane 2010). Climate model projections exhibit variability on decadal timescales, though they are initialized from randomly selected pre-industrial states, so variability in projections is not synchronized with observations (Meehl et al. 2009). Recent work has shown that it is theoretically possible to improve skill in predicting some aspects of global and regional climate over a decade by initializing climate models with observations, which has led to further research on decadal climate prediction, along with claims about its potential to anticipate climate impacts (Meehl et al. 2009, Murphy et al. 2010, Keenlyside and Ba 2010). However, it has not yet been demonstrated that decadal predictions have sufficient predictive skill to be used operationally.
Decadal climate modelling has evolved from seasonal climate modelling, which is able to provide useful forecasts on regional scales for temperature and precipitation in certain locations, particularly the tropics, and is regularly used operationally. On seasonal timescales predictability largely arises from slowly varying sea surface temperatures and major modes of variability such as the El Niño Southern Oscillation (ENSO) in the coupled climate system (Palmer and Hagedorn 2006). On decadal scales, predictability is believed to arise at least partly from lower-frequency climate modes and forced boundary conditions. Climate modes which may potentially offer predictability on decadal timescales include the Atlantic Multidecadal Oscillation, a basin-wide fluctuation of sea surface temperatures in the North Atlantic with a periodicity of around 70 years (Schlesinger and Ramankutty 1994) and the Pacific Decadal Oscillation, a pattern of climate variability in the Pacific (Mantua et al 1997). Predictability from boundary conditions comes from anthropogenic (e.g. greenhouse gas, aerosols) and natural (e.g. volcanic, solar) sources (Keenlyside and Ba 2010).

To explore the potential for skilful decadal prediction, the first multi-model decadal hindcast set was made as part of the ENSEMBLES project, covering the last 50 years (Van Der Linden and Mitchell 2009). The models used are state-of-the-art coupled ocean–atmosphere global circulation models, integrated forward for ten years from ten start dates distributed throughout the hindcast period. Previous work with this hindcast dataset has looked at skill in annual average temperature and precipitation for the first year from initialization and in four year average blocks thereafter (Oldenborgh et al 2012), focusing on North Atlantic and Pacific sea surface temperatures, and on global average temperature. However, climate impacts depend on time and space scales shorter than this; regional level sub-annual climate is the principal driver of climate impacts and interannual variations in seasonal temperature and precipitation over small regions can have significant socio-economic impacts, for example on agriculture and health as well as other sectors (Washington et al 2006). Therefore we focus here on using the ENSEMBLES hindcasts to answer the question: can the decadal hindcasts anticipate interannual variations in temperature and precipitation on the time and space scales relevant to climate impacts? We consider annual and seasonal averages, at different regional scales, to see if the predictions potentially available from these decadal hindcasts can drive impact models with skill. Details of models and validation methods are described in sections 2 and 3 contains results and discussion and conclusions can be found in section 4.

### 2. Methodology

The ENSEMBLES multi-model decadal stream 2 hindcasts consist of four forecast systems: IFS33r1, HadGEM2, ARPEGE4.6 and ECHAM5, developed at ECMWF, UK Met Office, CERFACS and IFM-GEOMAR respectively. All models include the main radiative forcings and none have flux adjustments at the ocean surface. Details of the models with further references can be found elsewhere (Oldenborgh et al 2012). Three members for each model were run for ten years starting on 1 November 1960, 1965 and every five years thereafter until 2005, giving nine hindcast time blocks and one which extends into the future (Van Der Linden and Mitchell 2009).

Throughout this analysis we look at annual (November to October) and seasonal averages, specifically boreal winter and summer (hereafter referred to as DJF and JJA respectively). To validate the hindcasts, multiple reference datasets have been used, details can be found in table 1.

Each of the models was found to have a temperature drift, dependent on time from initialization. To counter the confounding effect that this drift would have on the statistics, it was removed point-wise. To do this we first calculated the lead-time dependent drift by averaging all of the hindcasts from each of the four models separately to create four ten year time series. From these were subtracted the average of reference datasets averaged over the same periods (using NCEP when correlating against NCEP and ERA40 when correlating against ERA40), creating four drifts—one for each model, for every grid point. These lead-time dependent drifts were then subtracted from every member in the hindcasts corresponding to each model. Cross-validation was not used, and all subsequent analysis uses the drift-corrected data.

The similarity between hindcasts and observations at annual and seasonal scales was tested for by calculating Pearson’s product–moment correlation coefficients between the ensemble mean and each reference dataset across all lead times, for annual and for seasonal means. Two-tailed 99% significance levels were calculated based on Student’s $t$-test, dependent on sample size which varies with reference dataset. Furthermore, serial correlations were taken into account by adjusting the effective sample size according to:

$$n' \cong n \frac{1 - \rho_1}{1 + \rho_1},$$

where $n$ and $n'$ are the original and effective sample sizes respectively and $\rho_1$ is the lag-1 autocorrelation coefficient (Wilks 2006). Global maps of temperature correlations have been plotted, before and after detrending, along with precipitation correlations (detrended precipitation correlations

### Table 1. Summary of the reference datasets used in this study.

| Reference dataset | Dates       | Type                        | Reference                  |
|-------------------|-------------|-----------------------------|----------------------------|
| NCEP              | 1948–2010   | Reanalysis, 2 m temperature and precipitation | Kalnay et al (1996)        |
| ERA40             | 1950–2002   | Reanalysis, 2 m temperature and precipitation | Uppala et al (2005)        |
| GPCP              | 1979–2010   | Merged precipitation        | Adler et al (2003)         |
are not shown since the trend in precipitation was not considered to be significant). To detrend, we subtract the linear regression point-wise from each individual member averaged across all time blocks, doing the same for observations (i.e. separately for each block). We show only correlations with NCEP in this paper, commenting on results for the other reference datasets. Extra figures corresponding to the other reference datasets are available in the supplementary material (available at stacks.iop.org/ERL/7/044012/mmedia).

After assessing the skill of the ENSEMBLES decadal hindcasts in reproducing observed interannual precipitation and temperature there remains the question of the signal to noise ratio, namely to what extent predictable regional variations might rise above noise from uncertainties in the forced response of the simulated coupled climate system. Analysis of variance tests are generally employed to separate the total variability for a given climate variable into an unpredictable component (mainly arising from atmospheric dynamics and the ocean–atmosphere coupling at short timescales) and a potentially predictable component due to the slow varying external boundary forcing (anthropogenic such as greenhouse gases and natural such as volcanic). A one way analysis of variance (von Storch and Zwiers 1999) is here applied to the decadal ensemble hindcast to quantify the fraction of the variance of precipitation and temperature due to the common external forcing against the internal variability of the coupled atmosphere–ocean climate system. This ratio is also known as potential predictability (PP). A complete description of the mathematics describing the calculation of potential predictability is contained in the ESM.

For the decadal simulations, we concatenated all time blocks as a single time dimension before performing this analysis (to increase the sample size). We also calculated anomalies with respect to each the model ensemble mean before performing the analysis, to avoid deflating the PP due to the different nature of the model biases (3 members for 4 different models with potentially significantly different biases). PP is expressed in % and gives an idea about the magnitude of the predictable signal with respect to the total signal. This is estimated in the model world and does not provide any information about the realism of the atmospheric signal, which is why it is only useful alongside measures of skill, such as correlations.

The ability of the models to simulate trends over different regions was also explored. Drift-corrected hindcasts were averaged spatially over 18 regions and subsequently trends were calculated by a linear regression in the hindcasts and in the observations at each start date. Trends were calculated for each ensemble member separately. This was repeated for multiple trend lengths, from five up to ten years, each starting from the first year (i.e. year 1–5, 1–6, 1–7 etc) and was calculated for trends in both annual and seasonal averages. Subsequently, Spearman’s rank correlations between all the trends simulated by one model and one reference dataset were calculated, along with significance levels at the 90%, 95% and 99% level. In order to display the maximum amount of information possible only the exceedance of significance levels are indicated for each of the models, rather than providing exact value of the correlation.

3. Results

Figure 1 shows temperature and precipitation correlations between hindcasts and the NCEP reanalysis before and after detrending, for annual, DJF and JJA averages. Stippled areas indicate correlations significant at the 99% level.

Before detrending, the temperature correlations are generally significant globally at the 99% level, with correlations reaching around 0.6. Generally the hindcasts have slightly larger correlations when validated against NCEP than against ERA40 (though the spatial patterns are similar). Over land, the regions which have significant correlations at the annual level for both reference datasets are located around equatorial Africa, the Mediterranean region, Asia, South-West US and Greenland, with large areas of significant correlations over the north Atlantic and the Indian Ocean, with correlations of up to 0.6 (figure 1(a)). For seasonal averages the area of significant correlations is decreased, though with significant correlations across much of the globe, with maxima around the tropics (figures 1(d) and (g)).

After detrending the temperature data (figures 1(b), (e) and (h)), the correlations are not significant, suggesting that the significant correlations before detrending are caused by the long-term trend in temperature and beyond this any predictions of individual yearly or seasonal average temperature do not have skill.

Considering precipitation correlations (figures 1(c), (f) and (i)), whilst there are areas of significance for NCEP, there are almost no areas of significance for the other references used (ESM figure 3), meaning that the significant precipitation correlations shown in figure 1 are not robust. This suggests that the models are unable to make skilful predictions of individual annual or seasonal averages of precipitation.

Shown in figure 2 are maps of PP for temperature and precipitation. For temperature at the annual scale (figure 2(a)), there is a band of PP greater than 30% laying over the tropics, with a maximum of 60% over the maritime continent, with PP lower than 30% elsewhere. For seasonal averages (figures 2(c) and (e)) the PP maximum still lies over the tropics, though is generally not greater than 40%. This is consistent with studies showing that there is more predictability for temperature over the tropics than the extra tropics (Palmer and Hagedorn 2006). There is also a slight seasonal pattern to PP, with a northward (southward) movement of the maximum PP band in boreal summer (winter), though this shift is not pronounced. Precipitation PP is much lower than for temperature, which is also consistent with previous studies (Palmer and Hagedorn 2006). For annual averages over land it barely reaches greater than 5% (figure 2(b)). The maximum over the whole globe is 8%, over the Maritime Continent. This suggests that the spread of precipitation hindcasts over land is such that the potential for predictability may not be sufficient to be useful.

Shown in figure 3 are the regions considered for trend analysis, and figure 4 shows the significance levels for temperature and precipitation trend correlations for annual averages. For temperature at the annual scale, significant correlations for all models are observed for global land–sea
Figure 1. Correlation between the ensemble mean of the ENSEMBLES decadal hindcasts and NCEP reanalysis, for all lead times and start dates. Temperature correlations before (after) detrending are shown in the left (centre) column; precipitation correlations are shown in the right column. Precipitation data has not been detrended before correlating. Results for annual, DJF and JJA averages (for all lead times) are shown in the top, middle and bottom rows respectively. The stipple area represents areas of correlation significant at the 99% level. The greyed out area in the precipitation plots indicate regions where model precipitation climatology is less than 1 mm/day.

4. Discussion and conclusions

The analysis of the ENSEMBLES decadal hindcasts presented in this paper suggests that the prediction skill of the models for temperature is limited to annual global land–sea and global land trends, and there is no skill in precipitation predictions. There are some areas of significant correlation over the globe for temperature before detrending, though with trends removed the correlation between hindcasts and multiple reference datasets are below useful levels of significance. It has also been shown that the correlations and PP are lower for seasonal averages than for annual averages. Trend correlations are significant for globally averaged temperature for multiple trend lengths, and a few regions have been identified which have significant temperature correlations, independent of the reference dataset. Correlations for precipitation are only above significance for NCEP, and for the other reference datasets they are below significance everywhere, suggesting that they are not robustly significant. Precipitation also has low PP, and trend correlations are below significance. These results are consistent with other studies, which find the global average temperature trend represented well in these hindcasts, and do not find significant skill for precipitation for four year averages (Oldenborgh et al. 2012). This work then extends this conclusion to predictions at annual and seasonal scales.
Figure 2. Temperature (left column) and precipitation (right column) potential predictability for annual averages ((a) and (b)), DJF ((c) and (d)) and JJA ((e) and (f)). Greyed out area in the precipitation plots indicate regions where the ensemble mean climatology averaged across all models is less than 1 mm/day. Note the different scale for the colour bar in the temperature and precipitation plots.

Figure 3. Regions considered for trend correlation analysis.
Figure 4. Multi-year trend correlation significance levels for annually averaged temperature (left) and precipitation (right) (reference dataset is NCEP reanalysis). Each quadrant in each square stands for one of the four models in the ENSEMBLES decadal hindcasts (clockwise from top left: UK Met Office, ECMWF, IFM-GEOMAR, CERFACS). The three variations in warm (cold) colours indicate Spearman’s rank correlation coefficients significantly above (below) zero at the 90%/95%/99% levels respectively (levels at ±0.324, ±0.382, ±0.491).

Whilst there may be skill in prediction of near term evolution of global average temperature, this does not equate to the ability to say something useful about climate on the scale on which it impacts human society (Oreskes et al 2011). It is questionable how useful a skilful prediction of the annual global average temperature trend is if one is interested in making predictions of phenomena which unfold on regional, sub-anual scales. Annual global average temperature masks temporal and spatial variability, and it is variability which drives climate impacts (Washington et al 2006). Coupled with poor predictions of precipitation, this suggests that current decadal predictions are not skilful at impact scales and that forecasts made by them are not of sufficient quality to drive models of climate impacts (such as agricultural, health or hydrological models).

Even though there may be little useful skill in direct model output of temperature and precipitation it may still be possible to predict impacts on decadal timescales. This may be done by relating impacts to potentially predictable variables other than air temperature and precipitation (Smith et al 2010). It may also be possible to predict climate impacts using dynamical or statistical methods which predict the evolution of low-frequency oceanic oscillations such as the Atlantic Multidecadal Oscillation and the Pacific Decadal Oscillation (Enfield and Cid-Serrano 2006) and then relating them to climate impacts. However, there are several potential sources of uncertainty to consider when attempting to predict impacts in this indirect way: uncertainty in the exact nature of the teleconnection between a large scale climate mode and regional climate impact, uncertainty in the prediction of the oscillation itself, and uncertainty due to the unpredictability of the forcing, due to the state of higher-frequency climate modes such as ENSO which are high source of climate variability and may potentially interact with low-frequency modes in unknown ways.

There are some limitations to this study. The first is the limited number of start dates available in the hindcasts. Generally more validation points gives more confidence, and with only nine hindcast start dates confidence in validation is limited. A second limitation is the relatively small size of
the initial condition ensemble. Ensemble size is particularly important when estimating potential predictability, and it is questionable that three members per model is sufficient to robustly estimate this. A final caveat to note is that there is uncertainty in the reference datasets to which the hindcasts are compared, particularly for reanalysis datasets, which are not necessarily representative of reality. This is particularly the case in places where observations are sparse (e.g. in Africa in the 1990s).

Finally, it is important to stress that whilst the skill of the models is low, decadal prediction is still in its infancy. Furthermore, initialized climate models run in decadal mode have the potential to be useful in other ways. They have the advantage over uninitialized climate projections in that they can potentially predict internal climate variability, they can help to inform model development, are useful to learn about model biases, initialization strategies and climate variability. They can also help to build trust in climate projections. Nevertheless, we conclude here that the generation of decadal climate models used as part of the ENSEMBLES project have not demonstrated the ability to make useful predictions of climate at the scales at which it impacts society. This is a negative result, but it is nonetheless an important one for relevant communities to understand, as understanding current limitations of predictions allows efforts made to adapt to changes in climate to be focused wisely (Oreskes et al. 2011).

Decadal models will continue to be developed (for example in the current CMIP5 experiments), and their eventual role in climate change adaptation policy it is not yet clear. It is therefore important to maintain effective communication between research communities along the science-policy spectrum, such that research and policy expectations of decadal prediction remain informed by reality.

Acknowledgments

For invaluable discussion of work leading to this paper the authors would like to gratefully thank Francisco Doblas-Reyes and Lenny Smith and their colleagues at IC3 and LSE respectively, as well as the three anonymous reviewers whose comments have greatly improved the work. Dave MacLeod recognizes support from NERC for the PhD stipend and to the EU-FP6 ENSEMBLES project for model data. Andy Morse acknowledges NERC EQUIP funding and Cyril Caminade funding from EU-FP7 QWeCl.

References

Adler R F et al 2003 The version-2 global precipitation climatology project (GPCP) monthly precipitation analysis (1979–present) J. Hydrometeor. 4 1147–67
Cane M A 2010 Climate science: decadal predictions in demand Nature Geosci. 3 231–2
Enfield D B and Cid-Serrano L 2006 Projecting the risk of future climate shifts Int. J. Climatol. 26 885–95
Kalnay E et al 1996 The NCEP/NCAR 40-year reanalysis project Bull. Am. Meteorol. Soc. 77 437–71
Keenlyside N S and Ba J 2010 Prospects for decadal climate prediction Wiley Interdiscip. Rev.: Clim. Change 1 627–35
Mantha N J, Hare S R, Zhang Y, Wallace J M and Francis R C 1997 A pacific interdecadal climate oscillation with impacts on salmon production Bull. Am. Meteorol. Soc. 78 1069–79
Meehl G A et al 2009 Decadal prediction Bull. Am. Meteorol. Soc. 90 1467–85
Murphy J, Kattsov V, Keenlyside N, Kimoto M, Meehl G, Mehta V, Pohlmann H, Scaife A and Smith D 2010 Towards prediction of decadal climate variability and change Proc. Environ. Sci. 1 287–304
Oldenborgh G J, Doblas-Reyes F J, Wouters B and Hazeleger W 2012 Decadal prediction skill in a multi-model ensemble Clim. Dyn. 38 1263–80
Oreskes N, Stainforth D A and Smith L A 2011 Adaptation to global warming: do climate models tell us what we need to know? Philosophy Sci. 77 1012–28
Palmer T and Hagedorn R 2006 Predictability of Weather and Climate: From Theory to Practice (Cambridge: Cambridge University Press)
Schlesinger M E and Ramankutty N 1994 An oscillation in the global climate system of period 65–70 years Nature 367 723–6
Smith D M, Eade R, Dunstone N J, Fereday D, Murphy J M, Pohlmann H and Scaife A A 2010 Skillful multi-year predictions of Atlantic hurricane frequency Nature Geosci. 3 846–9
Uppala S M et al 2005 The ERA-40 re-analysis Q. J. R. Meteorol. Soc. 131 2961–3012
Van Der Linden P and Mitchell J F B 2009 ENSEMBLES: Climate Change and its Impacts: Summary of Research and Results from the ENSEMBLES Project Technical Report (Exeter: Met Office Hadley Centre) (online http://ensembles-eu.metoffice.com/docs/Ensembles_final_report_Nov09.pdf)
von Storch H and Zwiers F W 1999 Statistical Analysis in Climate Research vol 95 (Cambridge: Cambridge University Press)
Washington R et al 2006 African climate change: taking the shorter route Bull. Am. Meteorol. Soc. 87 1355–66
Wilks D S 2006 Statistical Methods in the Atmospheric Sciences (New York: Elsevier)