Multi-Year Predictability and Prediction

Timothy DelSole

George Mason University, Fairfax, Va and Center for Ocean-Land-Atmosphere Studies, Calverton, MD

September 19, 2012

In collaboration with Michael Tippett and Liwei Jia
Why Study Multi-year Prediction?
Why Study Multi-year Prediction?

Predictions are critical tests of our understanding of the climate system.
Forced Predictability

Models using both natural and anthropogenic forcing

Models using only natural forcings

IPCC-AR4, fig. 9.5
Evidence of Unforced Multi-year Predictability

Figure: Trajectories of leading principal component of 170m ocean temperature simulated by GFDL model; from Griffies and Bryan (1997).
Claims of Multi-Year Predictability in Stationary Models

- Latif and Barnett (1994, Science)
- Grötzner et al. (1999, J. Climate)
- Boer (2000, Climate Dyn.)
- Delworth and Mann (2000, Climate Dyn.)
- Collins (2002; Climate Dynamics)
- Collins and Sinha (2003; Geophys. Res. Letters)
- Pohlmann et al. (2004; J. Climate)
- Latif et al. (2004; J. Climate)
- Branstator et al. (2011; J. Climate)
Percent of Predictable Variance of Decadal Mean 2m-Temperature

Figure: Boer and Lambert (2008)
Challenges in Predicting the Real Climate System

- **Forced Predictability**
  - accurate predictions of anthropogenic and natural forcing
  - accurate predictions of the response to climate forcing.

- **Unforced Predictability**
  - accurate observations of the subsurface ocean
  - construct initial conditions without “shock” or “drift”
  - different models imply different levels of predictability
Dynamical Predictions (Keenlyside et al. 2008)

Figure: Correlation skill in predicting observed 10-year mean surface temperature anomalies a decade in advance for 9 ICs during 1955-2005.
CMIP5 Decadal Hindcast Experiments
Skill of CMIP5 Decadal Predictions

Fig. 2. Local anomaly correlation of (left) years 2-5 and (right) years 6-9 hindcasts for the CMIP5 models compared to damped persistence and the empirical multivariate AR1 model, for hindcasts initialized yearly from 1960-2000. (a) Damped persistence (b) empirical multivariate AR1 model (LIM) (c) HadCM3 (d) MPI-ESM-LR (e) GFDL-CM2p1. Contour interval is 0.1 with negative values indicated by blue shading. Shading of positive values starts at 0.1; redder shading denotes larger values of correlation.

from fig. 2 of Newman (2012)
Empirical Predictions

- Empirical predictions avoid many of the problems faced by dynamical models. For example:
  - They avoid “initialization shock” and “climate drift”.
  - They do not require a dynamically complete state vector.
- Empirical predictions are easier than dynamical predictions
- On decadal time scales, much predictability comes from slowly varying anomalies, which can be predicted by empirical models.
- Predictability of linear models can be diagnosed in ways that nonlinear models cannot.
Models in Empirical Decadal Prediction

- **persistence**
- **pointwise regression**
- **linear inverse model (LIM)**
- **multivariate linear regression**
- **constructed analogue**

\[
\begin{align*}
  x_{t+\tau} &= x_t \\
  x_{t+\tau} &= \beta_{\tau} x_t \\
  x_{t+\tau} &= L^\tau x_t \\
  x_{t+\tau} &= \sum_{n \in \text{training}} w_n x_{n+\tau}
\end{align*}
\]
Standard implementations of the constructed analogue method give forecasts that are the same as linear regression.
## Skill of Constructed Analogue for NINO3.4

<table>
<thead>
<tr>
<th>EOFs</th>
<th>10</th>
<th>25</th>
<th>30</th>
<th>35</th>
</tr>
</thead>
<tbody>
<tr>
<td>correlation (CV)</td>
<td>0.51</td>
<td>0.29</td>
<td>0.22</td>
<td>0.12</td>
</tr>
<tr>
<td>variance fraction (CV)</td>
<td>0.51</td>
<td>1.18</td>
<td>1.66</td>
<td>2.45</td>
</tr>
</tbody>
</table>

forecasts made in the beginning of July for the following March-May average (lead-8) of the NINO34 index during the period 1955-2003.
Predictions of Simulated SST (Hawkins et al. 2011)

It is clearly important to consider more than one measure of prediction skill when analysing prediction systems as these two metrics have produced different estimates of which method works best, and the choice of which to use will be situation dependent.

4.3 Where does the long lead time skill in HadGEM1 come from?

Although the lagged correlations predictions tend to perform best for shorter lead times, the methods which use non-local information tend to perform better for longer lead times, suggesting that they are predicting some of the dynamical evolution of the SSTs.

The tropical north Atlantic skill in HadGEM1 at long lead times is an intriguing region of skill to explain, as this region has low potential predictability (Fig. 1). Interestingly, the skill is larger for the mean of years 6–10 than the mean of years 1–5 (not shown) suggesting that the skill comes from a non-local source. To more convincingly demonstrate the non-local mechanisms, a series of data withholding experiments were performed. Figure 6 shows how the correlation skill for HadGEM1 for a lead time of 6–10 years changes as different regions are masked out of the construction of the statistical model—a far North Atlantic region (northwards of 47°N) and a Gulf Stream region (GSR). These regions are chosen as they are significantly correlated with the tropical north Atlantic at 6–10 year lead times (not shown).

For the Atlantic LIM method (left column), removing the far North Atlantic region completely removes the skill from the tropical north Atlantic. For the Atlantic CA method (right column), both regions seem important, and removing each in turn reduces the skill, which again disappears completely when both regions are removed from the domain. A final test, removing the tropical region itself (bottom row), shows that even when the local tropical data is not used in the construction of the statistical model, there

- $L$ estimated from 7 EOFs of 140 years pre-industrial control.
- CA estimated 140 years of PiCNTRL, all grid points
- 10-year prediction based on preceding 140 year training data.
- Predictability of Pacific SST limited to under two years.
Correlation skill is 1 (perfect!). MSE grows quadratically with time.
correlation skill of surface temperature for leads 6-9 years
- 8, 6, 6 EOFs of IndoPacific, Atlantic, land surface temperature
- $L$ from 1yr lagged 1900-2009 HadISST data (“observations”).
- Leave-10-out cross validation.
Diagnosis of LIM Predictability (Newman 2012)

\[ x_{t+\tau} = L^\tau x_t = \sum_{k} \lambda_k^\tau u_k \left( v_k^T x_t \right) \]
Least Damped Eigenmode of LIM (Newman 2012)

Fig. 5. Leading empirical normal modes, with their associated projection coefficient time series. Contour interval is the same in all panels. Sign is arbitrary but is consistent with coefficient time series. Red shading indicates one sign, and blue shading indicates the other sign.

“Secular trend pattern”
most energetic phase: Atlantic decadal variability
least energetic phase: PDO when ENSO is removed
Empirical Decadal Predictions

- Newman (2007)
- Kruger and von Storch (2010)
- Hawkins et al. (2011)
- Zanna (2012)
- Newman (2012)
- DelSole, Jia, and Tippett (2012)
DelSole, Jia, and Tippett (2012)

Fit $L_\tau$ to pre-industrial control runs, then predict observations.

- $L_\tau$ is determined independently of observations (no artificial skill).
- Lots of data available (1500 years for training and verification)
- Regression model has skill only to the extent that dynamical models capture the correct space-time structure of observed variability.
- Regression model fitted to 20 EOFs from 8 CMIP5 control runs
Predictions of Observed SST (DelSole, Jia, Tippett 2012)

Figure: Skill for observed **unforced** N. Atlantic SST during 1910-2004.

- **Red**: Predict independent twentieth century runs
- **Blue**: Predict independent pre-industrial control runs
- **Green**: Persistence
- **Black**: Predict observed **unforced** anomalies
Average Predictability Time (APT)

\[ \text{APT} = 2 \int_0^{\infty} \left( \frac{\sigma^2_{\text{clim}} - \sigma^2_{\text{forecast}}(\tau)}{\sigma^2_{\text{clim}}} \right) d\tau \]

DelSole and Tippett (2009a, 2009b)
Decomposing Predictability

Find components that maximize APT (DelSole and Tippett 2009).

\[ APT = 2 \int_0^\infty \left( \frac{\sigma^2_{clim} - \sigma^2_{forecast}(\tau)}{\sigma^2_{clim}} \right) d\tau \]
Most Predictable Component in CMIP5 Pre-industrial Control Runs

A

B

obs

AMO
Skill in Predicting the Most Predictable Component in Observations (1910-2004).
2nd and 4th Predictable Components

2nd PrC (PDO)  

4th PrC

![Maps of global climate patterns showing 2nd Predictable Component (PDO) and 4th Predictable Component.](image)

![Graphs showing skill development over lead time for observed (obs), 20C control, and persistence.](image)
Coupling between 2nd and 4th Predictable Components
Decadal Experiments with GFDL Model (Yang et al. 2012)

![SST IMP](image)

**a**

- Map showing SST IMP with contour lines indicating temperature variations.

**b**

- Graph showing time series of AMO index and ERSST with years ranging from 1940 to 2020.

- The AMO index is represented by a blue line, and ERSST is represented by a red line.

- The graph displays peaks and troughs indicating fluctuations in SST IMP over the years.
Can the IMP and forced response be distinguished?
Fingerprinting Method

Fit observed annual average SST to

\[ T_{obs}(x, y, t) = a_{for}(t)T_{for}(x, y) + a_{imp}(t)T_{imp}(x, y) + w(x, y, t) \]

**Detection:** Test hypothesis \( a_{for}(t) = 0 \).

**Attribution:** Test hypothesis \( a_{for}(t) = \) predicted amplitude.
Forced-to-Unforced Discriminant

Find the pattern $T_{for}(x, y)$ that maximizes detectability in models (Jia and DelSole 2012).
Forced Pattern

shaded area: 95% confidence interval of forced pattern in observations.

blue line: Ensemble mean amplitude of forced pattern in models
shaded area: 66% confidence interval of IMP in observations.

red line: Observed Atlantic Multidecadal Oscillation (AMO) index.

**Unforced climate models naturally simulate a multi-decadal component very similar to the AMO.**

---

DelSole, Tippett, and Shukla (2011)
Global Mean Sea Surface Temperature

Spatially Averaged SST on 'Well-Observed' Grid

Forced + IMP
Forced Only
Observation

Temperature Difference (K) from 1901–1950

year

1900 1920 1940 1960 1980 2000

−0.4 −0.2 0.0 0.2 0.4 0.6
Global Mean Sea Surface Temperature

Spatially Averaged SST on 'Well-Observed' Grid

-0.4 -0.2 0.0 0.2 0.4 0.6

1900 1920 1940 1960 1980 2000

Forced + IMP
Forced Only
Total
Model Mean

Temperature Difference (K) from 1901−1950

year
Both dynamical and empirical models can skillfully predict SSTs on multi-year time scales.

Linear models can be decomposed into components that explain skill.

Unforced climate models naturally simulate a multi-decadal component very similar to the AMO.

Empirical model derived from dynamical models show skill of predicting certain components of annual mean SST up to 9 years.

Recent decadal predictions show hindcast skill in N. Atlantic.