Predictability at the Intersection of Weather and Climate

Prashant Sardeshmukh

Climate Diagnostics Center, CIRES, University of Colorado
And Physical Sciences Division / ESRL / NOAA
Boulder, Colorado
Introduction and Motivation

1. How skillful are state-of-the-art coupled GCMs at making Week-2 and Week-3 forecasts in the Tropics and Extratropics?

   We compare the forecast skills of two CGCMs (NASA and NCEP) with the skill of simple Linear Inverse Models (LIMs), based on observed lag-covariances in the Tropics (of OLR) and Extratropics (of $\Psi_{200}$ and $\Psi_{850}$).

2. What is a realistic predictability estimate for this timescale?

   We estimate the expected anomaly correlation forecast skill using the CGCMs and LIMs under a “perfect model” assumption, and argue that the LIM-based predictability estimates are more accurate.

3. Are forecasts reaching a predictability limit for these timescales?

   We compare the GCMs’ actual forecast skill with our predictability estimates.
Introduction and Motivation continued....

The NCEP Medium range Forecast (MRF) model was unable to outperform a LIM for forecasts of 250hPa streamfunction at Week 3.

Do the new (and coupled) GCMs perform better?

From Newman, Sardeshmukh, Winkler and Whitaker 2003 (MWR)
# Models and Data

## CGCM Hindcasts used

<table>
<thead>
<tr>
<th>Model</th>
<th>AGCM</th>
<th>OGCM</th>
<th>Hindcasts</th>
<th>Initialization</th>
</tr>
</thead>
<tbody>
<tr>
<td>NASA/GEOS5</td>
<td>GEOS 5 2x2.5x72</td>
<td>MOM4 360x200x50</td>
<td>6-month hindcasts from 1980-2005</td>
<td>Daily 21z from replay runs</td>
</tr>
<tr>
<td>NCEP/CFS03</td>
<td>GFS T62L64</td>
<td>MOM3</td>
<td>9-month hindcasts from 1981-2005</td>
<td>15 times per month from R2 (atm) and GODAS (ocn)</td>
</tr>
</tbody>
</table>

## Reanalyses & OLR Datasets used

<table>
<thead>
<tr>
<th>Data Set</th>
<th>Years Used</th>
<th>Resolution (lon x lat)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NCEP/NCAR Reanalysis (R1)</td>
<td>1980-2008</td>
<td>2.5x2.5</td>
</tr>
<tr>
<td>NCEP/DOE Reanalysis (R2)</td>
<td>1980-2008</td>
<td>2.5x2.5</td>
</tr>
<tr>
<td>ERA40 Reanalysis</td>
<td>1980-2001</td>
<td>2.5x2.5</td>
</tr>
<tr>
<td>20th Century Reanalysis</td>
<td>1980-2005</td>
<td>2.5x2.5</td>
</tr>
<tr>
<td>MERRA Reanalysis</td>
<td>1980-2005</td>
<td>2/3x1/2</td>
</tr>
<tr>
<td>JRA25 Reanalysis</td>
<td>1980-2008</td>
<td>2.5x2.5</td>
</tr>
<tr>
<td>NOAA/AVHRR OLR</td>
<td>1980-2008</td>
<td>2.5x2.5</td>
</tr>
</tbody>
</table>
Linear Inverse Model (LIM)

General Description

Assume the system can be described as:

\[ \frac{dx}{dt} = Bx + \text{stochastic noise} \]

The solution is:

\[ x(t + \tau) = \exp(B\tau)x(t) + \varepsilon \]

The best forecast (in a least squares sense) is:

\[ x(t + \tau) = \exp(B\tau)x(t) \]

For more details see Penland and Sardeshmukh 1995 (J. Clim)
Linear Inverse Model (LIM)

Implementation of LIM

Define a reduced system anomaly state vector
\[
X = \begin{bmatrix} \psi \\ H \end{bmatrix}
\]
\(\psi = 200\text{mb and 850mb NH streamfunction anomalies for DJF (30 EOFs)}\)
\(H = \text{Tropical OLR anomalies for DJF (7 EOFs)}\)

In practice, \(X\) also needs to be coarse-grained & time averaged
1. Interpolate to a 7.5 x 3.5 deg grid.
2. Use 7-day running means.

Make Forecasts: \(x(t + \tau) = \exp(\mathbf{B}\tau)x(t)\)

1. Estimate \(\mathbf{B}\) (from observed and Reanalysis data)
2. Cross-Validate forecast skill through “jacknifing”

\[
B = \frac{1}{\tau} \log \left[ C(\tau) C^{-1}(0) \right]
\]

See Winkler, Newman, and Sardeshmukh 2001 (J. Clim) and Newman, Sardeshmukh, Winkler, and Whitaker 2003 (MWR) for details of method

From Pegion and Sardeshmukh 2011

The LIM is a skillful model that provides a good baseline for assessing the skill of the CGCMs
Comparison of Coupled Model Skill and LIM Skill

Difference in Anomaly Correlation Skill: CGCM skill minus LIM skill

Week-3 Forecasts

\[ \Psi_{200} \quad \Psi_{850} \]

GEOS5 minus LIM

CFS minus LIM

LIM remains a challenging benchmark for CGCMs to beat at Week-3 in the extratropics

Difference in Anomaly Correlation Skill: GEOS5 7-member Lagged Average Ensemble - LIM

Week-2

Week-3

\[ \Psi_{200} \quad \Psi_{850} \]

A “poor man’s” 7-member lagged-average ensemble forecast is made from the GEOS5 forecasts. The skill of the ensemble-mean forecast is generally higher than the LIM skill in the extratropics at both Week 2 and Week 3.
For all variables and models, anomaly correlation skill at these forecast ranges in generally very low (mostly < 0.5). This low skill limits the utility of the forecasts. What are the prospects for skill improvement? To address this, it would be nice to have estimates of the potential skill.
Taylor Diagram Summary of Coupled GCM Skill and LIM Skill

For all variables and models, anomaly correlation skill at these forecast ranges in generally very low (mostly < 0.5). This low skill limits the utility of the forecasts. What are the prospects for skill improvement? To address this, it would be nice to have estimates of the potential skill.

Red Symbols : PSI 200  
Blue Symbols : PSI 850  
Green Symbols : OLR
For all variables and models, anomaly correlation skill at these forecast ranges in generally very low (mostly < 0.5). This low skill limits the utility of the forecasts. What are the prospects for skill improvement? To address this, it would be nice to have estimates of the potential skill.
Predictability ( = “Potential Skill”)

The “potential skill” is the expected skill of infinite-member ensemble-mean forecasts using a “perfect model”

Signal-to-noise ratio:

\[ S = \frac{\|\text{ensemble mean anomaly}\|}{\|\text{ensemble spread}\|} \]

Expected Anomaly correlation skill of a Perfect Model, n-member ensemble:

\[ \rho_n(\tau) = \frac{S^2(\tau)}{\left\{ S^2(\tau) + 1 + \frac{1}{n} \right\}^{1/2}} \]

See Sardeshmukh, Compo and Penland. 2000 (J. Clim) & Compo and Sardeshmukh 2004 (J. Clim) for details
Estimating the Signal-to-Noise ratio $S$

For GEOS5:

$S$ is calculated directly from the 7-member lagged ensembles

For LIM:

$x(t) = \exp(Bt) \ x(0) + \text{noise} = G(t)x(0) + \text{noise}$

The variance is partitioned into the part due to the signal and the part due to the noise

\[
F = \text{diag} \left[ G(\tau)C(0)G^T(\tau) \right] = \text{Signal Variance}
\]

\[
E = \text{diag} \left[ c(0) - G(\tau)C(0)G^T(\tau) \right] = \text{Noise Variance}
\]

\[
S^2 = \frac{F}{E}
\]
GEOS5 ensemble-based predictability estimates are ~ 0.3-0.4 higher than the estimates made from the LIM. Is this realistic?
The noise variance from the GEOS5 ensemble is much smaller than the forecast error variance, indicating that it is underestimated. However, the LIM noise variance is similar to its forecast error variance.

The GEOS5 autocorrelations are generally higher than in the Reanalysis, indicating that the model is more persistent than observed, leading to an overestimated signal.
How much Predictability is left to be realized?

Difference Between GEOS5 Lagged Ensemble Skill and LIM-based Predictability Estimates

Week-2

Week-3

In the extratropics, there is little skill left to be realized.
Conclusions

1. An ensemble forecasting system is needed for the CGCMs to outperform the LIM at Week-3.
2. Traditional “perfect” model predictability estimates using a CGCM overestimate the magnitude of the forecast signal and underestimate the magnitude noise leading to inflated predictability estimates.
3. The LIM provides a more realistic “lower upper bound” for potential skill than the lagged average ensemble from GEOS5.
4. For NH streamfunction and central Pacific OLR, there is apparently little skill left to be realized on average.

Forecasts on these timescales should be focused on making “forecasts of opportunity”, i.e. when the forecast signal is identified a priori to be relatively large
State-Dependent Predictability
The most rapidly growing singular vectors of the LIM operator $G(t) = \exp(Bt)$ can help identify relatively more skillful forecast cases \textit{a priori} . . .

Expected and actual pattern correlation skill of Week-3 N.H. forecasts, stratified by initial state projections on the right singular vectors of $G$ ($t=21$ days)

For wintertime Week 3 Forecasts of 250 mb streamfunction

From Newman et al 2003
Another example of the usefulness of LIM: diagnosis of coupled interactions

Observed Power spectra of the leading Tropical SST and Atmospheric Diabatic Heating EOFs (red curves), compared to spectra predicted by the Coupled-LIM (blue curves) and by the Uncoupled-LIM (green curves).

Gray shading represents 95% confidence intervals determined from a 2400 yr run of the C-LIM.

Insets in each panel show the corresponding EOF and the variance of weekly anomalies explained by that pattern.

Dashed curves: spectra of the observed heating PC 1 projected onto the subset of either the ”Coupled” (yellow) or “Internal” (pink) eigenmodes of the full LIM operator.

From Newman, Sardeshmukh and Penland (J. Climate 2009)
Thank You