Decadal prediction: pain and potential

Wm. Merryfield, S. Kharin, G.J. Boer, W-S Lee, ... Canadian Centre for Climate Modelling and Analysis Environment Canada
Topics

- Potential predictability and skill
- Realizing the potential
  - Initialization and integrations
  - Bias correction
  - Some basic results
- What have we learned
Predictability Studies

Diagnostic
- analog methods
- potential predictability

Prognostic
- perfect model
- weather/climate forecasting
Predictability and prediction

- **predictability**
  - a characteristic of a *physical system* itself
  - a measure of the *rate of separation* of *initially close* states
  - indicates the *possibility of prediction*

- **prognostic predictability studies**
  - typically use a model to simulate “rate of separation”
  - presumption that the model “similar enough” to real system
  - model is not perfect but code characterizes it “perfectly”

- **forecast skill**
  - characterized by *error growth* rate, decorrelation, or other measure (rate of separation of actual and predicted states)
  - indicates the current *ability to predict*

- **potential predictability**
  - analysis of variance measure
  - identifies regions where *long timescale variance* is a useful fraction of the total
  - meant to indicate that prediction is *potentially possible*
  - *upper limit* of skill (if known from observations)
How do we determine the predictability and forecast skill of the system on decadal timescales?

- **Prognostic perfect model predictability studies**
  - Griffies and Bryan (1997)
  - Boer (2000)
  - Collins (2002)
  - Collins et al. (2006)
  - Branstator et al. (2010)
  - and others

- **Diagnostic potential predictability studies**
  - Pohlmann et al. (2004)
  - Predicate (2004...)
  - Boer and Lambert (2008)
  - and others

- **Investigations of forecast skill**
  - Smith et al. (2008)
  - Keenlyside et al. (2008)
  - Pohlmann et al. (2009)
  - CMIP5/IPCC (2013)
Climate Change 2013
The physical and biophysical basis

and decadal prediction
Decadal predictability and prediction

- potential predictability $p$
  - skill measures and $p$
  - multi-model estimates of $p$
- CCCma DHFP (decadal historical forecasting project)
  - approach
  - “bias correction”
  - forecast skill
- implications
Statistics of the *internally generated* component

- Model control runs - no external forcing
- Variable X is expressed as
  \[ X = \mu + \chi + x \]
  - \( \mu \) is the mean
  - \( \chi \) is the long timescale *internally generated* component, presumed to be at least partially predictable
  - \( x \) is the short timescale "noise" component which is unpredictable on the timescale considered
- Associated variances are
  \[ \sigma^2_X = \sigma^2_\chi + \sigma^2_x \]
- Potential predictability variance fraction (ppvf) is
  \[ p = \frac{\sigma^2_\chi}{\sigma^2_X} = \frac{\sigma^2_\chi}{(\sigma^2_\chi + \sigma^2_x)} \]
Potential predictability statistics

- Internally generated variability
  - Observations: \( X = \mu + \chi + x \)
  - Forecasts: \( Y = \nu + \psi + y \)

- Statistics
  - variances:
    \[
    \sigma^2_X = \sigma^2_\chi + \sigma^2_x \\
    \sigma^2_Y = \sigma^2_\psi + \sigma^2_y \Rightarrow \sigma^2_\psi + \sigma^2_y / N \text{ for ensemble mean}
    \]
  - potential predictability variance fractions (ppvfr)
    \[
    p_\chi = \frac{\sigma^2_\chi}{\sigma^2_X} \text{ for observations} \\
    p_\psi = \frac{\sigma^2_\psi}{\sigma^2_Y} \text{ for model} \\
    \Rightarrow 1 \text{ with } \sigma^2_y / N \Rightarrow 0 \text{ for large ensemble-average}
    \]
  - covariances:
    \[
    \text{cov}(XY) = \sigma_\chi \sigma_\psi R_{\chi\psi}(\tau)
    \]
    for \( R_{\chi\psi}(\tau) \) the correlation of \( \chi, \psi \) at forecast range \( \tau \)
Connection of $p$ with skill

- **Correlation skill**
  \[ r_{XY}(\tau) = \frac{\text{cov}(XY)}{\sigma_X \sigma_Y} = \frac{\sigma_\chi \sigma_\psi R_{\chi\psi}(\tau)}{\sigma_X \sigma_Y} = \left(p_\chi p_\psi\right)^{1/2} R_{\chi\psi}(\tau) \]
  - for “good model” ($\sigma_\chi = \sigma_\psi$, $\sigma_X = \sigma_Y$) single forecast
  \[ \Rightarrow p_\chi^{1/2} R_{\chi\psi}(\tau) \]
  - large ensemble-mean forecast ($\sigma^2_Y/N \Rightarrow 0$, $p_\psi \Rightarrow 1$)
  \[ \Rightarrow p_\chi^{1/2} \] for a “perfect” model with ($p_\psi, R_{\chi\psi} = 1$)

- **Other skill measures also depend on $p$**
  - **Mean square error** $m$
    \[ m_{XY}(\tau) = \sigma^2_\chi + \sigma^2_\psi - 2 \sigma_\chi \sigma_\psi R_{\chi\psi}(\tau) + (\sigma^2_X + \sigma^2_Y) \]
    \[ \Rightarrow 2\sigma^2_\chi (1 - p_\chi R_{\chi\psi}(\tau)) \] for good model single forecast
  - **Mean square skill score** $M$
    \[ M_{XY}(\tau) = \Rightarrow 2p_\chi R_{\chi\psi}(\tau) - 1 \] for good model single forecast
Multi-model *estimate* of *internally generated* long timescale potential predictability $p_\chi$

\[ X(t) = X_{\alpha^0} + (X_{\alpha^j} - X_{\alpha^0}) \]

\[ \sigma^2 = \sigma^2_\chi + \sigma^2_x \]

*unbiased* estimates

statistical *stationarity*
Temperature: potential predictability of *internally generated* variability $p_\chi = \sigma_\chi^2 / \sigma^2$ (%) for *decadal means* (CMIP3 multi-model control runs)

- Ratio of long timescale to total variance
- MME provides stability of statistics: $ppv_f$ in white areas <2% and/or not significant at 98% level
- Long timescale predictability found mainly over oceans
- Some incursion into land areas but modest $ppv_f$ (*denominator* is large)

Boer and Lambert, 2008
Long-timescale prognostic predictability study

- Initial “perfect model” predictability estimates from CCCma model
- Only 3-member ensemble

Decadal **perfect model** predictability
Cumulative "perfect model" predictability $p > 0.4$ at year 10

Boer, 2000: Clim Dyn
years 1-5

years 6-10

perfect model
predictability measure

North Atlantic

Pohlmann et al. 2004
**Precipitation**: potential predictability of internally generated variability \( p_\chi = \frac{\sigma^2_\chi}{\sigma^2} (%) \) for decadal means

- MME provides “some” significant areas of precipitation
- Much less potentially predictable than temperature
- Little incursion into land areas
- Precipitation predictability a weakened version of temperature predictability at these timescales
21st Century decadal potential predictability

- Variable now has forced component
  \[ X = \mu + \Theta + \chi + x \]
  with associated variances
  \[ \sigma^2 = \sigma^2_\Theta + \sigma^2_\chi + \sigma^2_x \]
  - \( \Theta \) is long timescale externally forced variability
  - \( \chi \) is long timescale internally generated variability
  - \( x \) is short timescale unpredictable “noise” variability
    - statistics pooled across models
- Potential predictability variance fraction now has two components
  \[ p = (\sigma^2_\Theta + \sigma^2_\chi)/\sigma^2 = p_\Theta + p_\chi \]
Forced plus internally generated case

○ Variability
  ● Observations:  $X = \mu + \Theta + \chi + \epsilon$
  ● Forecasts:  $Y = \nu + \Phi + \psi + \eta$

○ Statistics:
  ● variances:
    $\sigma^2_X = \sigma^2_\Theta + \sigma^2_\chi + \sigma^2_\epsilon$
    $\sigma^2_Y = \sigma^2_\Phi + \sigma^2_\psi + \sigma^2_\eta$
  ● potential predictability for forced and internally generated components
    $p_X = (\sigma^2_\Theta + \sigma^2_\chi) / \sigma^2_X = p_\Theta + p_\chi$ for observations
    $p_Y = (\sigma^2_\Phi + \sigma^2_\psi) / \sigma^2_Y = p_\Phi + p_\psi$ for model
  ● Correlation skill
    $r_{XY}(\tau) = \frac{\sigma_\Theta \sigma_\Phi R_{\Theta\Phi}(\tau)}{\sigma_\chi \sigma_\psi} + \frac{\sigma_\chi \sigma_\psi R_{\chi\psi}(\tau)}{\sigma_\chi \sigma_\psi}$
    $= (p_\Theta p_\Phi)^{1/2} R_{\Theta\Phi}(\tau) + (p_\chi p_\psi)^{1/2} R_{\chi\psi}(\tau)$

  $\Rightarrow p_\Theta^{1/2} R_{\Theta\Phi}(\tau) + p_\chi^{1/2} R_{\chi\psi}(\tau)$ for “good model, large-ensemble” forecast

○ forced contribution favoured
  ● $R_{\Theta\Phi} = 1$ for linear trend
  ● $\sigma^2_\Theta$ large hence $p_\Theta$ large for skill over a long forecast period (1961-2010)
  ● applies also to forecasts
21\textsuperscript{st} century temperature at a point
- forced component from 1\textsuperscript{st} decade

\[ \sigma_1^2 = \sigma^2_{\Theta_1} + \sigma^2_{\chi} + \sigma^2_x \]

\[ \sigma_\Delta^2 = \sigma^2_{\Delta\Theta} + \sigma^2_{\chi} + \sigma^2_x \]

\[ X = \mu + \Theta + \chi + x \]
MME (CMIP) next decade potential predictability of Temperature for the 21st century

- \( p_\Theta = \sigma^2_\Theta / \sigma^2 \)
  - *forced component* large over land
  - but *discounted* because noise is large

- \( p_\chi = \sigma^2_\chi / \sigma^2 \)
  - *internally generated component*
  - largest over high latitude oceans

- \( p = p_\Theta + p_\chi \)
  - *low, mid-latitude land*
  - *oceans excluding equatorial Pacific*

- In light of the potential predictability:
  - what are current “forecast skill” levels?
  - do they reflect the “potential predictability”?
Decadal predictability and prediction

- potential predictability
  - skill measures and $p$
  - multi-model estimates of $p$

- CCCma’s DHFP (decadal historical forecasting project)
  - approach
  - “bias correction”
  - forecast skill

- implications
CCCma (seamless) Predictions

CHFP2, DHFP1

Monthly-seasonal forecasts

45-60 days

Seasonal-annual forecasts

12 months

Annual-decadal forecasts

10-30 years

Global Land-Atmosphere Coupling Experiment (GLACE-2)

CanCM3

US Clivar Intraseasonal Prediction Experiment

WCRP CHFP

CHFP1
CHFP2a
CHFP2b

F. Lienert PDO
A. Ravindran MJO
G. Flato S. Kharin Sea ice

CMIP5 (IPCC AR5)

CanCM4

Climate projections to 2100+ (IPCC)

Merryfield
Kharin
Boer
Lee
Scinocca

.....
CanCM4 model components

- AGCM and OGCM new since AR4

- **CanAM4**
  - T63/L35
  - new shallow convection, radiation, aerosols...
  - includes “natural” volcanic & solar forcings

- **CanOM4**
  - 1.41°×0.94°× L40 (Δz~10m near surface)
  - GM stirring, KPP + tidal vertical mixing, anisotropic viscosity
  - solar penetrative heating according to climatological chlorophyll

- Earth-system version CanESM2 used for long-term AR5 simulations
## Decadal Historical Forecasting Project

<table>
<thead>
<tr>
<th></th>
<th>DHFP1A</th>
<th>DHFP1B</th>
<th>DHFP1C</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Model</strong></td>
<td></td>
<td>CanCM4 (AGCM4+OGCM4)</td>
<td></td>
</tr>
<tr>
<td><strong>Initialization</strong></td>
<td></td>
<td>IRU/CIN assim</td>
<td>+2Dvar assim with S-correction (anomaly)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Nudge to SST</td>
<td>+2Dvar assim with S-correction (full field)</td>
</tr>
<tr>
<td></td>
<td>atmos</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ocean</td>
<td>Nudge to SST</td>
<td></td>
</tr>
<tr>
<td></td>
<td>sea ice</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Ensemble members</strong></td>
<td></td>
<td>10</td>
<td></td>
</tr>
<tr>
<td><strong>Commencing dates</strong></td>
<td></td>
<td>1&lt;sup&gt;st&lt;/sup&gt; Jan each year</td>
<td></td>
</tr>
<tr>
<td><strong>Forecast duration</strong></td>
<td></td>
<td>10 years</td>
<td></td>
</tr>
<tr>
<td><strong>Retrospective period</strong></td>
<td></td>
<td>1961 (every 5 years) 2010 (completed for 1a, 1b, 1c)</td>
<td>1979 (every year) 2010 (1b only, completed)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1961 (every year) 1978 (1b only, in progress)</td>
<td></td>
</tr>
</tbody>
</table>

**DHFP1N**: ensemble of 10 uninitalized “freecasts”
DHFP initialization

AGCM CIN assim (ERA)
SST nudging (OISST)
Sea ice nudging (HadISST)

+ Anthropogenic forcing

1 Jan

Forecast

Asim1

Asim2

Asim10

Forecast 1

Forecast 2

Forecast 10

3D ocean T, S assimilation:
GODAS after 1981
SODA before:
DHFP1a: no sub-surface data
DHFP1b: full-field ocean data
DHFP1c: anomaly ocean data

multiple assimilation runs
**Incremental Reanalysis Update (IRU) assimilation:**

- run model freely for 3h ("forecast")
- difference with reanalysis → "centered" increments $\Delta x^a$
- rewind, rerun for 6h, adding analysis increments as forcing to model equations:

$$\frac{dx}{dt} = M(x) + h(t)\Delta x^a$$

* To better reflect observational uncertainties in ensemble, “dial back” assimilation *constant incremental nudging (CIN)*
Benefits of IRU/CIN

- accurate AGCM initialization
  - essential for 1st month skill
- ensemble generation
- better land initialization
- better ocean initialization/background state for assimilation

Due to “seeing” atmospheric forcing leading up to forecast
Impacts of AGCM assimilation on land initialization

Correlation of assimilation run vs Guelph offline analysis

Soil temperature (top layer)

Soil moisture (top layer)
Impacts of AGCM assimilation on ocean initialization vs nudging to SST

Correlations vs obs in equatorial Pacific (5S→5N)
Some initial results

- *drift correction*
- results for
  - temperature
  - sea ice
  - MOC
Drift/bias correction

- models “drift” from the initial state toward model climate which differs from the observed
  - $Y(t_j, \tau)$ the forecasts with $t_j$ the start times and $\tau$ the forecast range and $X(t_j, \tau)$ the corresponding observations
- the average “drift” is
  - $\Delta(\tau) = Y(t, \tau) - X(t, \tau)$ where the “dot” is the average over all the available start times $t_j$
- however, there are difficulties if there is a lack of stationarity in which case $\Delta = \Delta(t_j, \tau)$
Drift/bias correction and trend

Global mean temperature

“standard” anomaly bias correction introduces systematic behaviour when model and observed trends are different

(S. Kharin)
Bias removal vs trend adjustment

Bias-adjusted decadal hindcasts

Trend-adjusted decadal hindcasts
Drift/bias correction

- Generally need “correction” to make predictions useful (as in SIP)
- Non-stationarity/lack of cases adds difficulties
  - desirable to correct for “trend” but lack of data likely precludes this locally
  - serious risk of overfitting if you do anything other than basic bias correction
    - this can make “corrected” CMIP5 results untrustworthy and perhaps misleading
  - very little data involved in the standard CMIP5 case:
    - for forecasts every 5 years only 10 cases
    - ensembles and multi-model approaches don’t help – still only 10 cases
    - forecasts every year would help (although they are not independent)
Some initial results for Temperature

basic bias correction
10 cases
initial loss of skill for monthly, seasonal, annual averages

subsequent increase in skill for longer averages from forced component

initialized forecasts have improved skill compared to uninitialized for averages of a month to ~3-4 years

initial loss of skill for monthly, seasonal, annual averages

subsequent increase in skill for longer averages from forced component

initialized forecasts have improved skill compared to uninitialized for averages of a month to ~3-4 years

\( r^{1/2} \sim p \) measure of “potential” correlation

Temperature: global mean of correlation coefficient

(S. Kharin)
Temperature: global mean of correlation

- improvement of initialized forecasts over uninitialized freecasts for **25 cases** compared to CMIP5’s **10 cases**
- also get better scores verifying against GISS1200
Correlation score for temperature vs multi-model $p$

(only roughly related to the internal component)

<table>
<thead>
<tr>
<th>$\delta r^2$</th>
<th>initialized-uninitialized</th>
</tr>
</thead>
</table>

| $r^2$ | forced + internally generated |

basic bias correction
10 cases

(some agreement between $r^2$ and $p$)
MOC

1a: no subsurface ocean data

MOC estimates
Grist et al. JClim 2009
MOC

- ensemble of uninitialized simulations lack MOC variation
- surface forcing introduces MOC variation (but how realistic?)
  - MOC subsequently well predicted
- initialization with ocean information affects result, perhaps detrimentally

**Diagram:**
- **DHFP1A ANN MAX MERIDIONAL V PSI NATL 26N(SV) annual means**
- **CANESM2 ANN MAX MERIDIONAL V PSI NATL 26N(SV) annual means**
- **DHFP1B ANN MAX MERIDIONAL V PSI NATL 26N(SV) annual means**

**Legend:**
- adaptation runs
- ensemble mean
- +/-2 stand dev.

**Free running model**
- initialization uses subsurface ocean information
North Atlantic Overturning

Control run 10 year mean vs assimilation & forecast 1981 means (1st year of forecast)
Sea-ice
Summary

- Potential predictability measures give an indication of skill limits.
- Skill from forced component apparent; skill from initialized/internally generated component generally “weak”
  - skill mainly for temperature
  - perhaps sea ice, MOC
- General lack of cases a problem
  - only 10 cases for CMIP5 forecasts every 5 years
  - multi-model approaches can increase ensemble size but not cases
  - need decadal forecasts every year at minimum (underway)
- Bias correction and skill assessment aspects
  - non-stationarity a problem
  - need more cases for stable statistics
  - statistical approaches need great care
end of presentation
Drift/bias correction

- Generally need “correction” to make predictions useful (as in SIP)
- Non-stationarity adds difficulties
  - desirable to correct for “trend” but lack of data likely precludes this locally
  - serious risk of overfitting if you do anything other than basic bias correction
    - this can make “corrected” CMIP5 results untrustworthy and perhaps misleading
- Very little data involved in the standard CMIP5 case:
  - for forecasts every 5 years only 10 cases
  - ensembles and multi-model approaches don’t help – still only 10 cases
  - forecasts every year would help (although they are not independent)
- Implications
  - need forecasts initiated every year for better statistical stability
  - only “basic” bias corrected CMIP5 results made available?
  - what of bias correction for RCM driving fields?
  - statistical aspects in general need considerable care
Results depend on number of forecasts and verification

- better scores for:
  - larger sample (predictions every year)
  - GISS as reference data (smoother?)