Selecting appropriate bias correction schemes for seasonal-to-decadal prediction

Bias Corrections in Subseasonal to Interannual Predictions
Virtual Workshop

Phil Sansom and Chris Ferro
Exeter Climate Systems, University of Exeter, UK
E-mail: p.g.sansom@exeter.ac.uk

In collaboration with
Lisa Goddard and Simon Mason
IRI, Columbia University, USA
Introduction

1 Project outline

2 Comparing post-processing methods and hindcast designs

3 A family of statistical post-processing methods

4 Preliminary results from CMIP5

5 Conclusions
Objectives:

1. Develop a robust methodology for recalibrating model predictions;
   - Linear adjustments to mean and variance of forecast distribution.

2. Quantify how forecast performance depends on the design of prediction systems;
   - Length and frequency of hindcast record;
   - Number of ensemble members;
   - Effect of initialization.

3. Recalibrate existing ensemble experiments and determine the efficacy of their designs.
   - ENSEMBLES models;
   - CMIP5 models.
Different perspectives, similar problems

Given an existing ensemble forecast system and accompanying hindcast experiment, what is the optimal method of post-processing in order to obtain more skillful forecasts?

What is the optimal design for a new ensemble forecast system and accompanying hindcast experiment, in order to obtain more skillful forecasts after post-processing?
Comparing forecast performance

What other factors do we need to consider when comparing post-processing methods or hindcast designs?

- Natural variability;
- Time varying predictability;
- Observation quality.
Comparing forecast performance

What other factors do we need to consider when comparing post-processing methods or hindcast designs?

- Natural variability;

  Compare average performance over many forecasts.

- Time varying predictability;

- Observation quality.
Comparing forecast performance

What other factors do we need to consider when comparing post-processing methods or hindcast designs?

- Natural variability;
  Compare average performance over many forecasts.

- Time varying predictability;
  Compare performance based on forecasting same events.

- Observation quality.
Comparing forecast performance

What other factors do we need to consider when comparing post-processing methods or hindcast designs?

- Natural variability;
  
  Compare average performance over many forecasts.

- Time varying predictability;
  
  Compare performance based on forecasting same events.

- Observation quality.
  
  Compare performance based on recent forecast performance.
Comparing forecast performance

Should we use cross-validation or out-of-sample prediction when comparing post-processing methods or hindcast designs?

Cross-validation:
+ efficient design;
+ limits influence of changes in observation quality or predictability;
− may not be representative of future performance.

Out-of-sample prediction:
+ representative of real-world performance;
− inefficient design;
− sensitive to changes in observation quality and predictability.
How might we process ensemble forecasts?

One family of statistical post-processing methods is defined by

$$Y \sim N \left( a + b\overline{X}, c^2 + d^2 S^2 \right)$$

$Y$ is the observed climate, and $\overline{X}$ & $S^2$ are the ensemble mean & variance.

The parameters are easily interpreted as:

- $a$ unconditional bias in the ensemble mean;
- $b$ conditional bias in the ensemble mean;
- $c^2$ unconditional bias in the ensemble variance;
- $d^2$ conditional bias in the ensemble variance.

How might we process ensemble forecasts?

One family of statistical post-processing methods is defined by

\[ Y \sim N\left( a + b\bar{X}, c^2 + d^2S^2 \right) \]

\( Y \) is the observed climate, and \( \bar{X} \) \& \( S^2 \) are the ensemble mean \& variance.

Special cases include:

- **a0c0** the climatological forecast \((b = 0, d = 0)\);
- **0101** the raw ensemble forecast \((a = 0, b = 1, c = 0, d = 1)\);
- **a1c0** additive bias correction with forecast uncertainty equal to the MSE of the corrected forecasts \((b = 1, d = 0)\);
- **abc0** linear bias correction with forecast uncertainty equal to the MSE of the corrected forecasts \((d = 0)\).

Initial results based on CCCma CanCM4 and UKMO HadCM3 models

- 10 ensemble members
- initialized every year from 1960–2009
- predict near surface (2m) temperature
- average temperature over years 1, 2–5, 6–9 and 2–9
- verify against HadCRUT3v (1850–2014)
- only consider grid boxes with no missing observations
- leave-one-out cross-validation
- continuous ranked probability score
- average score over hindcasts initialized 1995–2004
- compare fitting periods from 10–45 years
- fit by maximum likelihood
- mean score over all grid boxes as summary measure
How should we estimate forecast uncertainty?

The MSE of the adjusted forecasts outperforms other methods of estimating the forecast uncertainty.
How should we estimate forecast uncertainty?

The relationship between the different methods is consistent over time.
How should we estimate forecast uncertainty?

The MSE of the adjusted forecasts tends to outperform the scaled ensemble variance.
Linear bias correction outperforms additive bias correction.
The relationship between linear and additive bias correction is also consistent over time.
Linear bias correction tends to increase forecast skill over land.
How long should the fitting period be?

The optimum fitting period is between 15 and 25 years at short lead times.
How long should the fitting period be?

Skill compared to 30 year climatology: Year 1, Linear bias correction

The optimum fitting period is also consistent over time.
Are forecasts skillful at longer lead times?

Post-processed forecasts outperform climatological forecasts at short lead times,...
Are forecasts skillful at longer lead times?

...but at longer lead times climatological forecasts may outperform additive bias correction.
Are forecasts skillful at longer lead times?

Skill compared to 30 year climatology: Years 2–5, 10 year fitting period

Climatological forecasts consistently outperform additive bias correction at longer lead times.
Are forecasts skillful at longer lead times?

Skill at longer lead times is limited to particular regions.
Summary of preliminary results

For time averaged near surface (2m) temperature:

- the MSE of the post-processed forecasts outperforms methods of estimating the forecast uncertainty based on the ensemble variance;

- linear bias correction outperforms additive bias correction over land;

- 15-25 year fitting periods are optimal short lead times;

- post-processed forecasts outperform climatological forecasts at short lead times;

- post-processed forecasts outperform climatological forecasts in some regions at longer lead times.
Always compare post-processing methods or hindcast experiments in a paired design, i.e., by forecasting the same events;

Compare performance based only on recent hindcasts;

Check skill compared to short climatologies, not just 30 years;

Use a cross-validation approach where possible.
Further work

- Test significance of differences between methods and fitting periods;
- Examine performance of post-processed forecasts with very short fitting periods at lead times $\geq 1$ year;
- Investigate the effect of hindcast frequency;
- Extend the analysis to seasonal time scales;
- Investigate the effect of ensemble size.