WCRP REPORT
World Climate Research Programme

Data and bias correction for decadal climate predictions

January 2011

ICPO Publication Series no.150

CLIVAR is a component of the World Climate Research Programme (WCRP). WCRP is sponsored by the World Meterorological Organisation, the International Council for Science and the Intergovernmental Oceanographic Commission of UNESCO. The scientific planning and development of CLIVAR is under the guidance of the JSC Scientific Steering Group for CLIVAR assisted by the CLIVAR International Project Office. The Joint Scientific Committee (JSC) is the main body of WMO-ICSU-IOC formulating overall WCRP scientific concepts.

Bibliographic Citation

INTERNATIONAL CLIVAR PROJECT OFFICE, 2011: Decadal and bias correction for decadal climate predictions . January. International CLIVAR Project Office, CLIVAR Publication Series No..150 (not peer reviewed).

Data and bias correction for decadal climate predictions

1. Introduction

Because climate models are not perfect simulated climatologies will differ somewhat from observed climatologies. The model state will "drift" toward the model climate as the forecast progresses and this drift will be confounded with the climate evolution that is being predicted. For this reason, near-term climate predictions are usually "bias corrected".

2. Forecast and climatological averages

We consider a sequence of "raw" climate forecasts $Y_{j\tau}$ where $j = 1, \ldots n$ identifies the initial times and $\tau = 1, \ldots m$ the forecast range. Most commonly in the climate context the $Y_{j\tau}$ will be monthly, seasonal, annual or multi-annual means of the quantity of interest such as temperature or precipitation. Observation-based data used to initialize and verify the climate forecasts are represented as $X_{j\tau}$.

The averages over a sequence of forecasts and the corresponding observations at forecast range range τ are

$\overline{Y}_\tau = \frac{1}{n}\sum_{j=1}^{n} Y_{j\tau}$

$\overline{X}_\tau = \frac{1}{n}\sum_{j=1}^{n} X_{j\tau}$

This *forecast average* will depend, of course, on the collection of forecasts that is available. For the Decadal Climate Prediction (DCP) component of CMIP5 this will generally encompass time periods from 1960-2010 but may differ from modelling group to modelling group.

The forecast averages are typically not the same as the climatological averages of the modelled and observed variables which will typically be averages from data sampled more frequently and possibly over longer time periods. Long-term climatological averages are represented here as $<Y>$ and $<X>$.

3. Initialization

There are two main approaches to forecast initialization referred to as "full field" and "anomaly" initialization. Bias correction and verification can depend on which initialization method is used.

Full field initialization constrains the model initial conditions to be close to the observed values with

$Y_{j\tau=0} \approx X_{j\tau=0}$

while *anomaly initialization* adds the observed anomaly to the model climatology of the variable in question as

$Y_{j\tau=0} \approx <Y> + (X_{j\tau=0} - <X>)$

In the case of an ensemble of forecasts the initial conditions are not all the same although "close" to one another. The nature of the perturbations around the value at $\tau = 0$ will depend on the approach taken by a particular modelling group.

4. Bias correction

a. For full field initialization

The bias correction is calculated from the collection of hindcasts and the corresponding observations. The model drift is estimated as the difference between *ensemble mean* forecasts and the observations averaged over all cases as

$d_\tau = \overline{Y}_\tau - \overline{X}_\tau$

and the *bias corrected* forecast $\hat{Y}_{j\tau}$ is obtained by subtracting this drift from each ensemble member as

$\hat{Y}_{j\tau} = Y_{j\tau} - d_\tau = \overline{X}_\tau + (Y_{j\tau} - \overline{Y}_\tau) = \overline{X}_\tau + Y'_{j\tau}$

Here $Y'_{j\tau} = Y_{j\tau} - \overline{Y}_\tau$ is the *anomaly* of the raw forecast with respect to the *forecast average.* It is generally considered to be best practice to calculate the bias correction in a "cross-validated" manner where the particular forecast to be corrected does not contribute to the forecast average.

b. For anomaly initialization

In this case the bias has been at least partially removed by the initialization method. The drift is characterized as the difference between the modelled and observed climates as

$D = <Y> - <X>$

and the bias corrected forecast is

$\hat{Y}_{j\tau} = Y_{j\tau} - D = <X> + (Y_{j\tau} - <Y>) = <X> + Y''_{j\tau}$

Note: that the "full field" version of bias correction can also be applied to anomaly initialized forecasts but may not be optimum.

5. CMIP5 decadal climate prediction data submission and data use

a. Raw data

Raw decadal prediction data sets as specified in documentation available from the CMIP5 website (http://cmip-pcmdi.llnl.gov/index.html) should be submitted in all cases. Users should regard these data as the basis for analysis.

b. Bias correction

Bias correction can be undertaken in order to use the decadal prediction results. The methods outlined in Sections 4a or 4b are recommended. In every case it is important to specify clearly what has been done when reporting results.

c. Bias corrected data

Data providers may also contribute a set of *bias corrected* CMIP5 data but this is limited by practicalities to basic climatological parameters. These comprise *monthly means* of bias corrected

- near-surface air temperature (tas)
- surface temperature (ts)
- precipitation rate (pr)
- sea level pressure (psl)

on the atmospheric grid.

If bias corrected data is supplied it is *necessary* that it be accompanied by a clear description of the bias correction method used and:

- for method 4a, the observational data set used must be be specified
- for method 4b, the observational $<X>$ and model $<Y>$ climatologies used must be specified

This information should be included in the "history" attribute attached to each of the bias corrected variables. In addition there is a request to collect this information from modelling groups using the so-called "METAFOR questionnaire".