Bias in decadal climate model forecasts

P. Geoffrey Allen
University of Massachusetts Amherst
Robert Fildes and Nikolaos Kourentzes
University of Lancaster

International Symposium on Forecasting
Santander, Spain
June 19-22, 2016

We gratefully acknowledge the assistance of Lenny Smith and Erica Thompson of the Centre for the Analysis of Time Series, London School of Economics, UK

Goals of the study

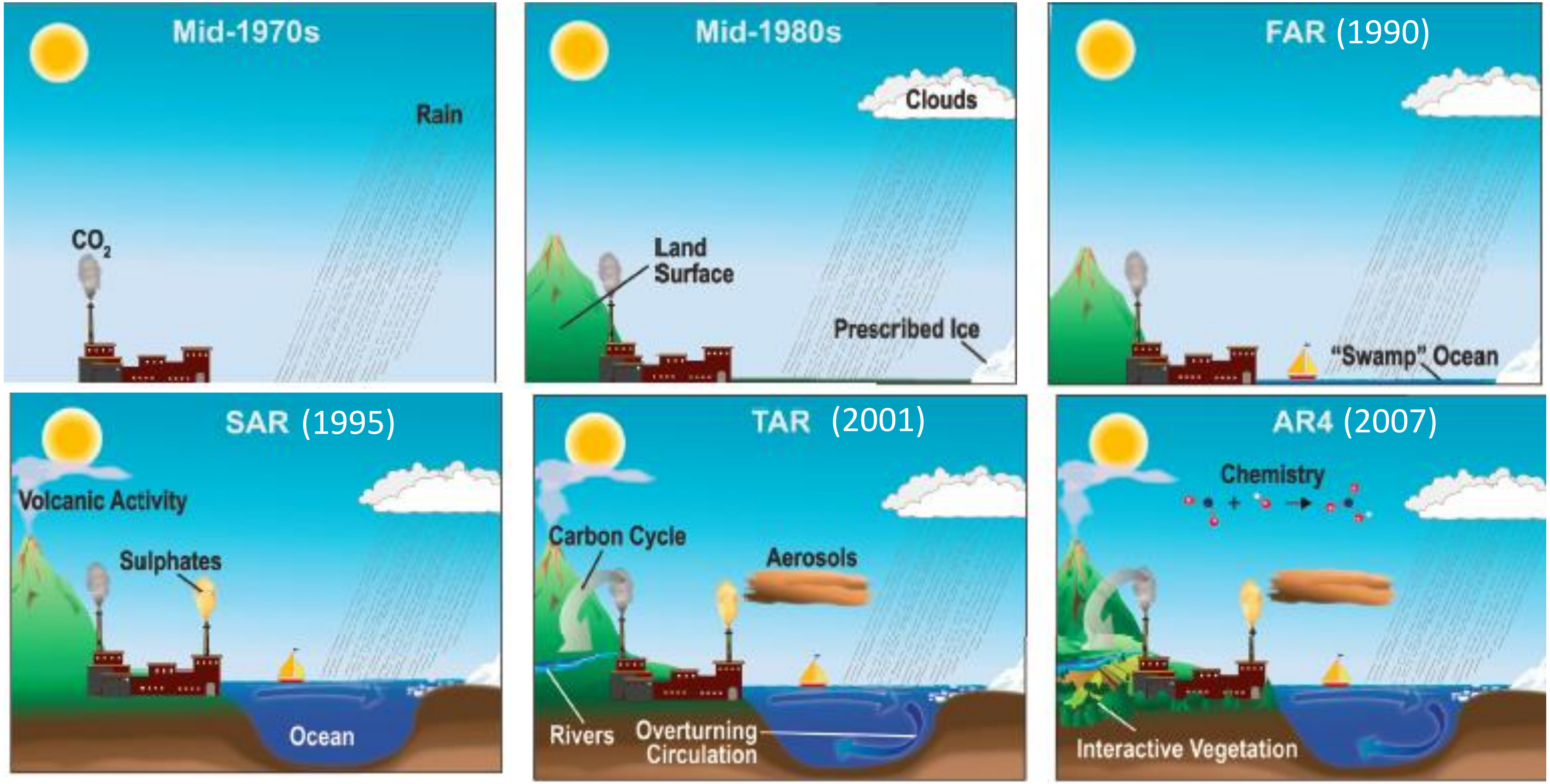
Measure bias in raw decadal forecasts (“hindcasts”) from Global Climate Models (GCMs)

De-bias raw decadal forecasts

Compare debiased GCM forecasts with simple statistical forecasts

Modern GCMs are complex . . .

Recap



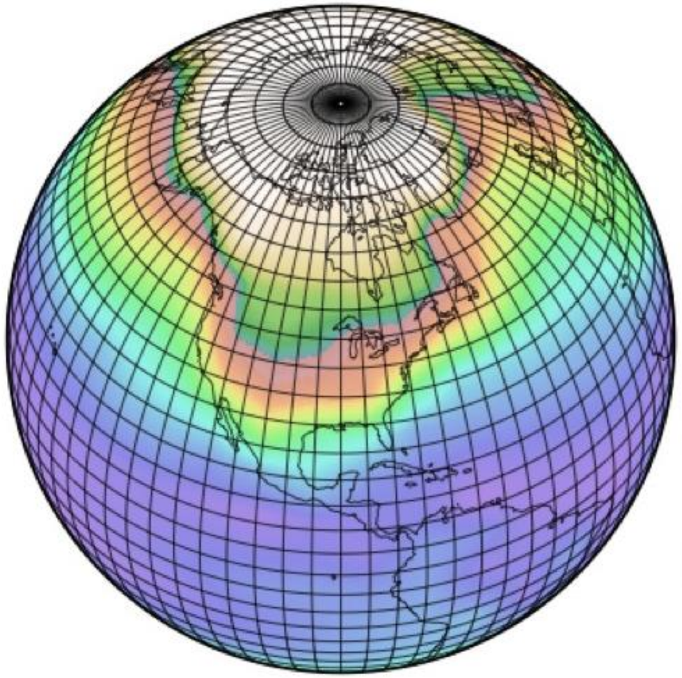
. . . but share common features

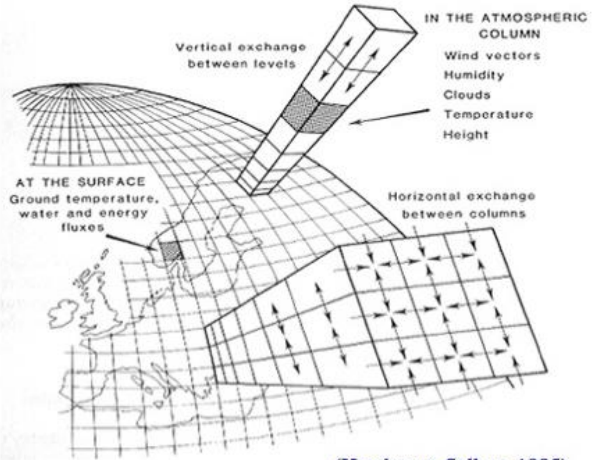
Recap

Their base is the "primitive equations" proposed by Bjerknes in 1904: partial differential equations (that must be approximated as difference equations for computation) for air movement, pressure, temperature and water content.

Vilhelm Bjerknes

They are grid point models, e.g., equal angle rectangles, the set-up for the observational series, HadCrut4





(Henderson-Sellers, 1985)

Sources of bias - 1

Recap

Models must be parameterized but this is done by "tuning" not by an algorithm that produces (at least in theory) unbiased estimators

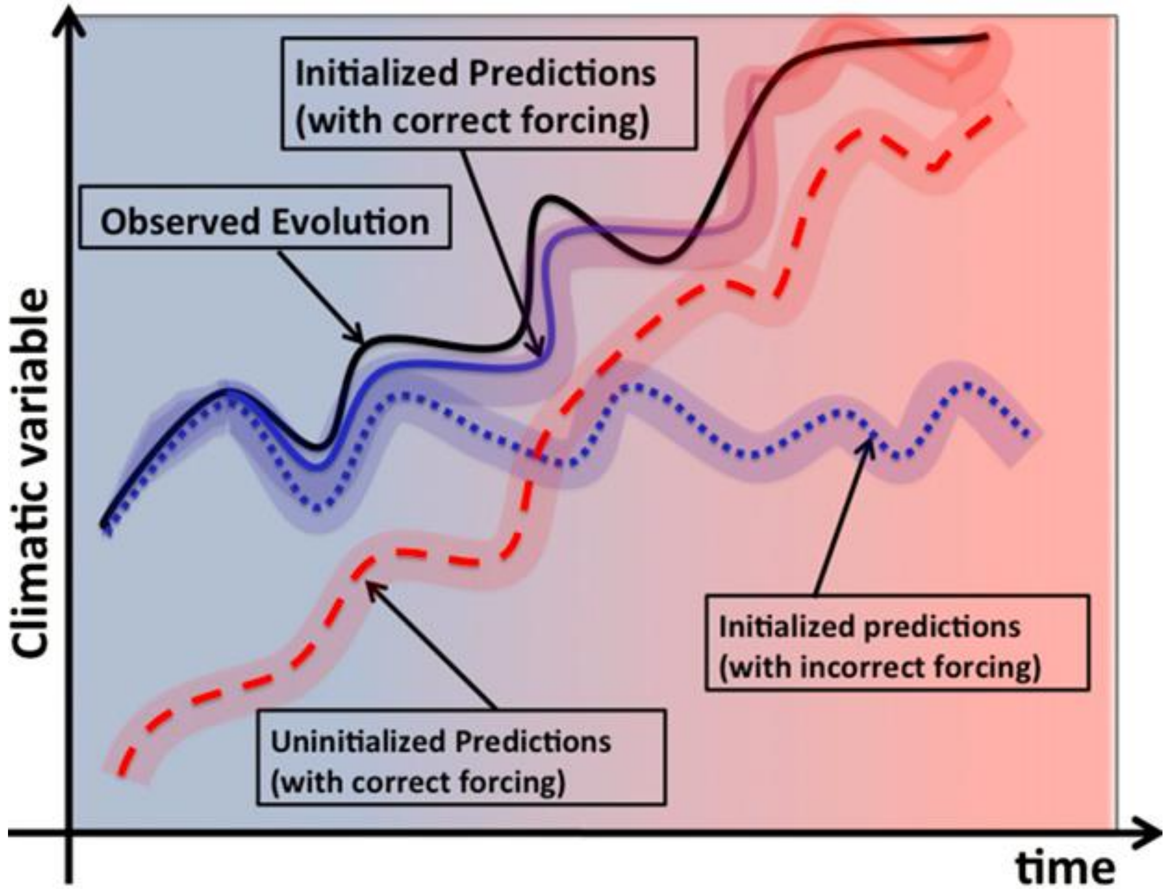
"The experiment contains empirical elements in that the representation of certain physical effects is based on meteorological experience with the actual atmosphere, rather than being predicted from the laws of physics."

Phillips, N. A. (1956). The general circulation of the atmosphere: A numerical experiment. *Quarterly Journal of the Royal Meteorological Society*, *82*(354), 123-164.

Norman Phillips
Constructor of the first computerized GCM

Sources of bias - 2

Recap



Decadal forecasts are started by either

- **Anomaly initialization** (the dashed red line) or
- **Full-field initialization**, from actual meteorological conditions (the blue lines)

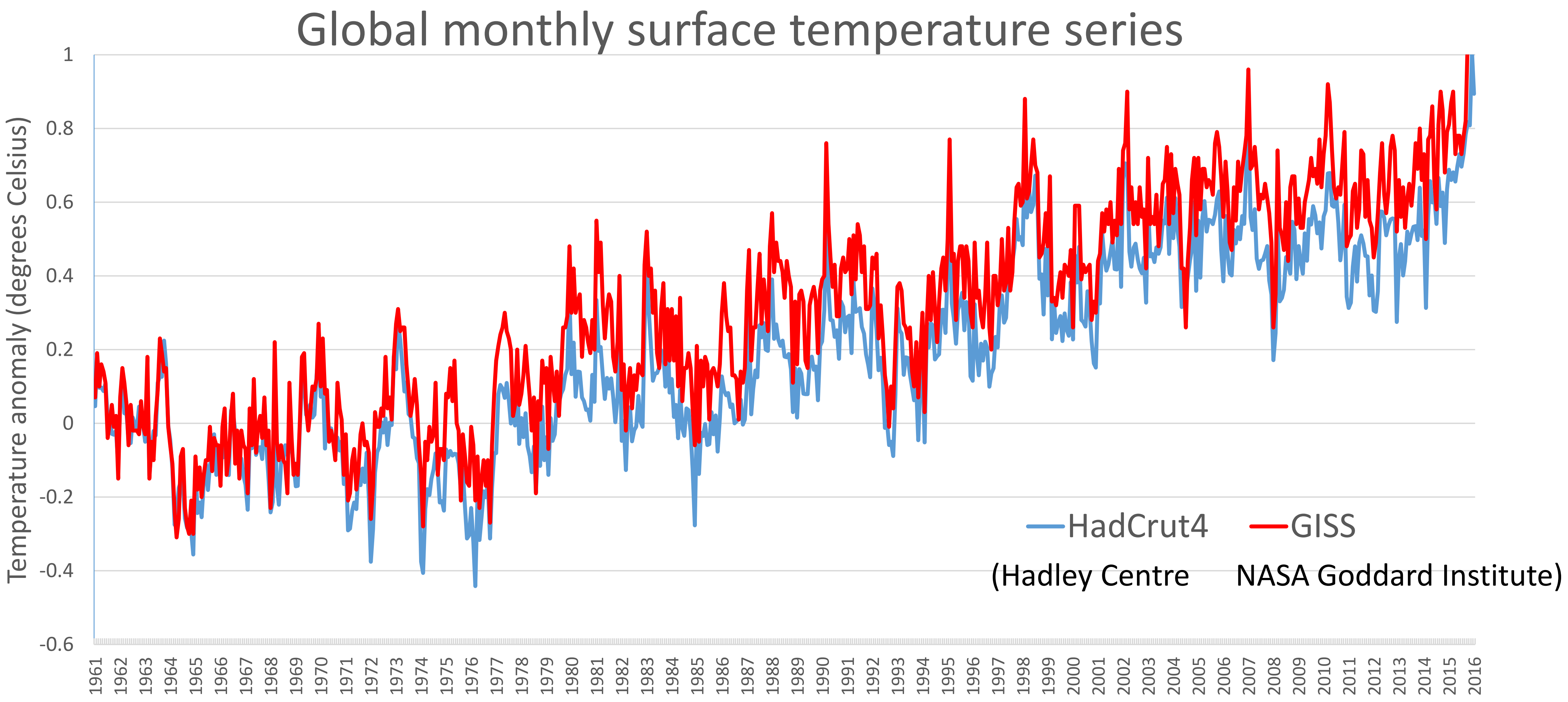
Source: Corti, S., Palmer, T., Balmaseda, M., Weisheimer, A., Drijfhout, S., Dunstone, N., ... & Storch, J. S. V. (2015). Impact of Initial Conditions versus External Forcing in Decadal Climate Predictions: A Sensitivity Experiment*. *Journal of Climate*, *28*(11), 4454-4470.

Observed series

Recap

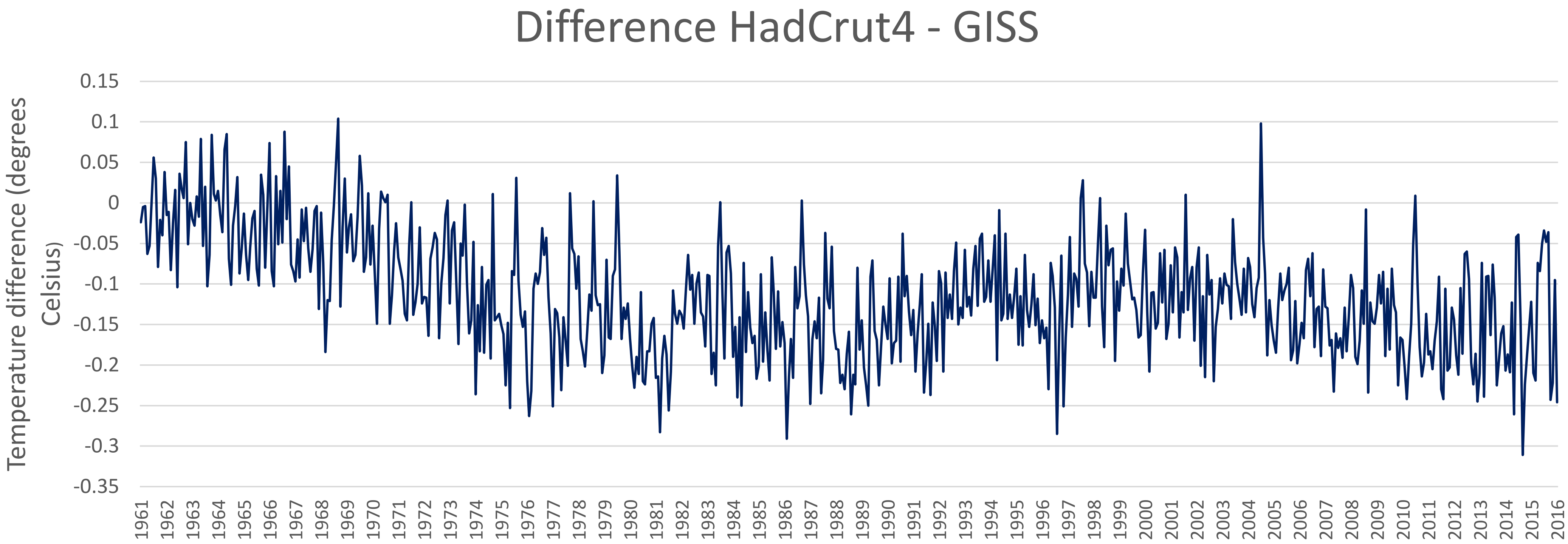
- Each weather station has its own climatology – the average temperature over 30 years, in this case 1961-90
- For climate work, temperatures are reported as *anomalies* or differences from the station's climatology
- Data reanalysis takes station anomalies to produce grid-point anomalies which are averaged over one month to produce a gridpoint monthly anomaly
- A weighted average over all grid points is the global average anomaly (in degrees Celsius)
- This must be related to GCM outputs in degrees Kelvin

There are several global surface temperature series. They are generally similar.



Differences occur, both between series and between updates of a series as a result of

- Different coverage (number of weather stations included)
- How *inhomogeneities* are dealt with (observations dropped, adjustments to observations)



Global surface temperature (GLO) 1961-90 climatology. It has annual variation of about 4˚C

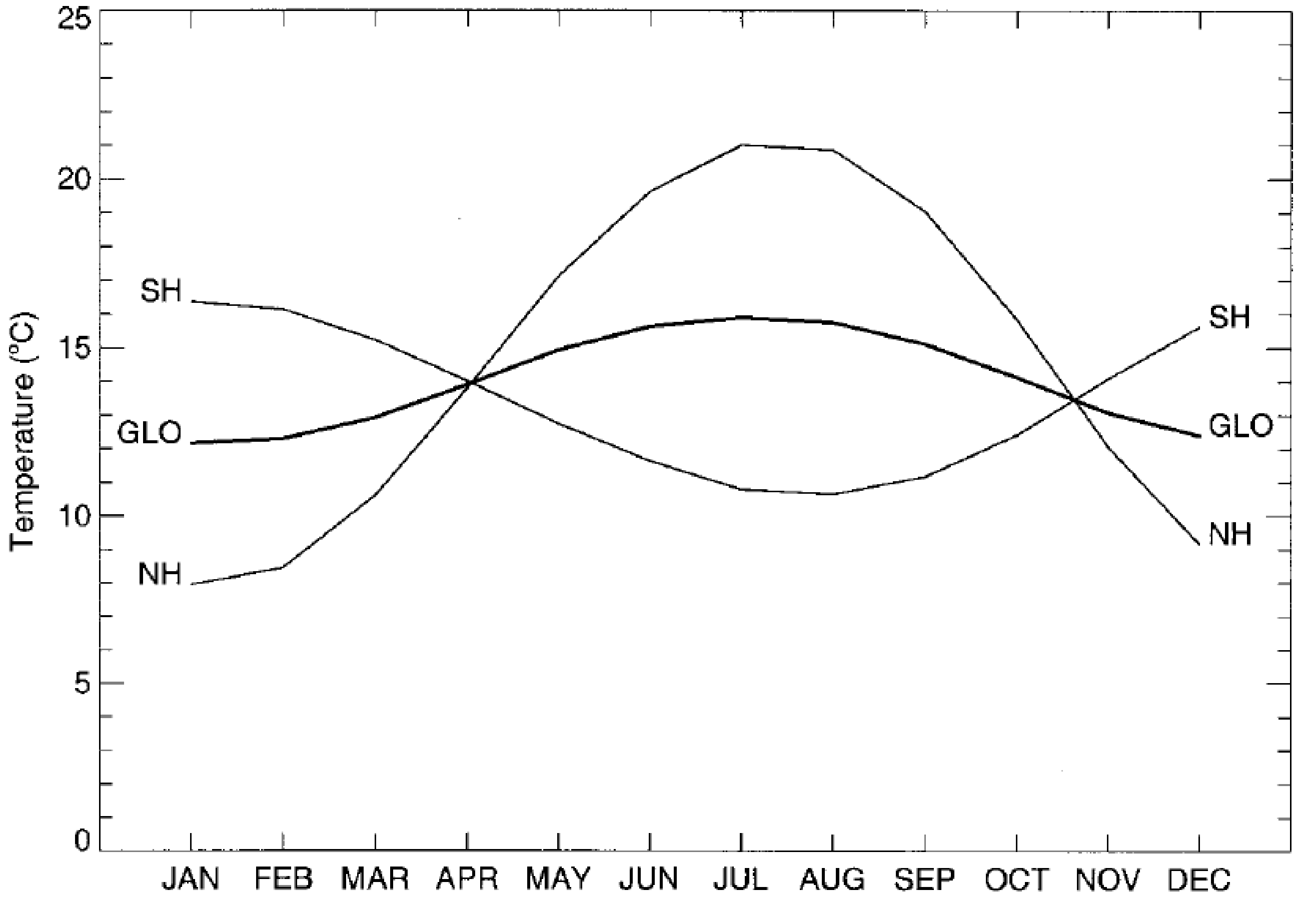


Figure 7. Seasonal cycle of hemispheric and global mean temperatures in absolute degrees Celsius based on the 1961–1990 period.

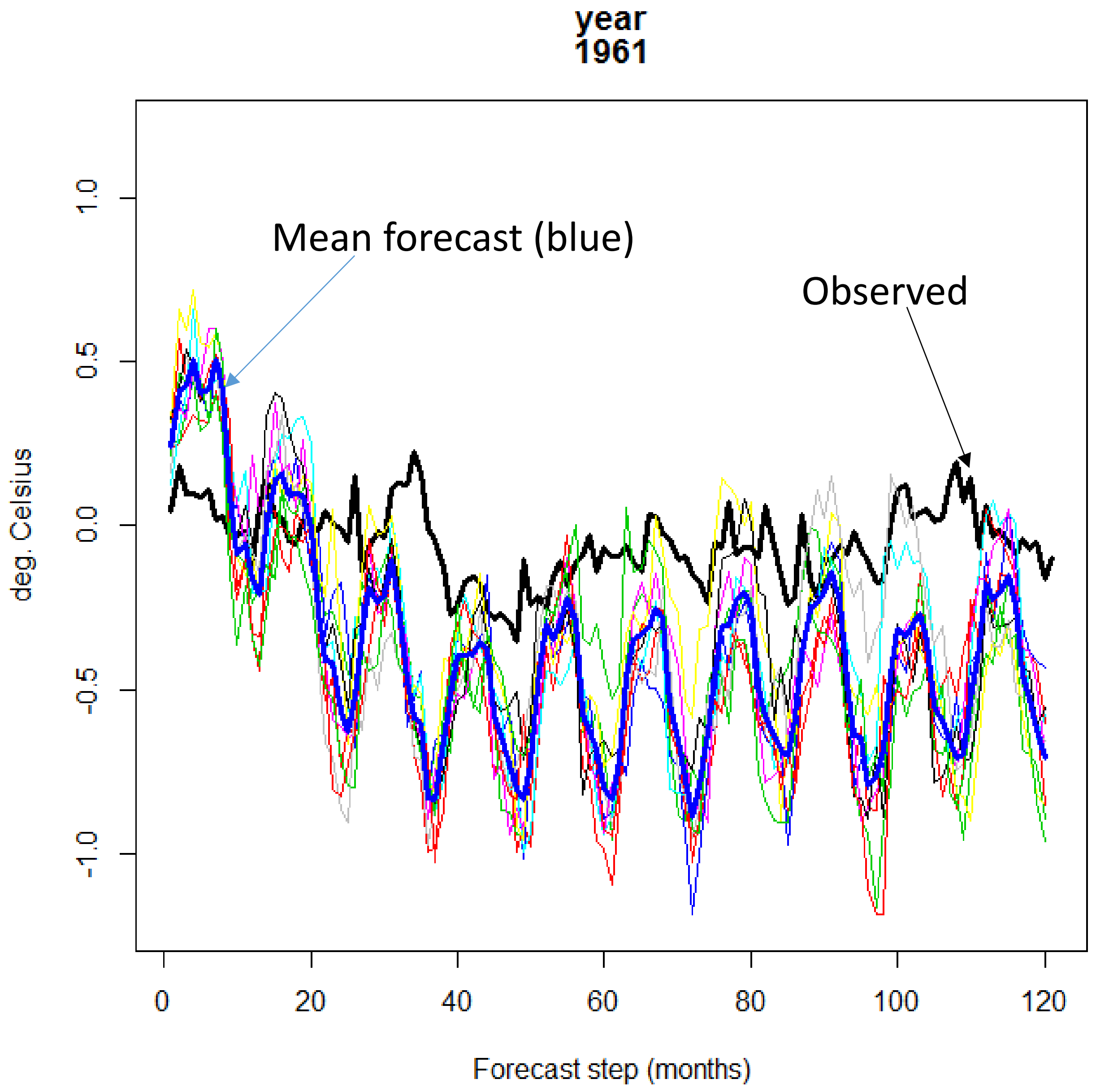
Source: Jones, P. D., New, M., Parker, D. E., Martin, S., & Rigor, I. G. (1999). Surface air temperature and its changes over the past 150 years. *Reviews of Geophysics*, *37*(2), 173-199.

Sources of Hindcast Data

- From the Fifth Coupled Model Intercomparison Project (CMIP5)
  - 10-year hindcasts every 5 years starting in November or December 1960
  - 6 series from 4 different models
  - Ensemble of 10 runs, for each series for each forecast origin
  - Monthly average from simulations of (usually) 6-hour time step averaged over all grid points
  - Temperatures reported in degrees Kelvin

Example of decadal forecast (CanCM4i1) converted to anomalies, 10 ensemble runs for 1961 start date. Black line: HadCrut4 observations.

All models and years show a similar seasonal pattern. The forecasts have a bigger amplitude than the observations.



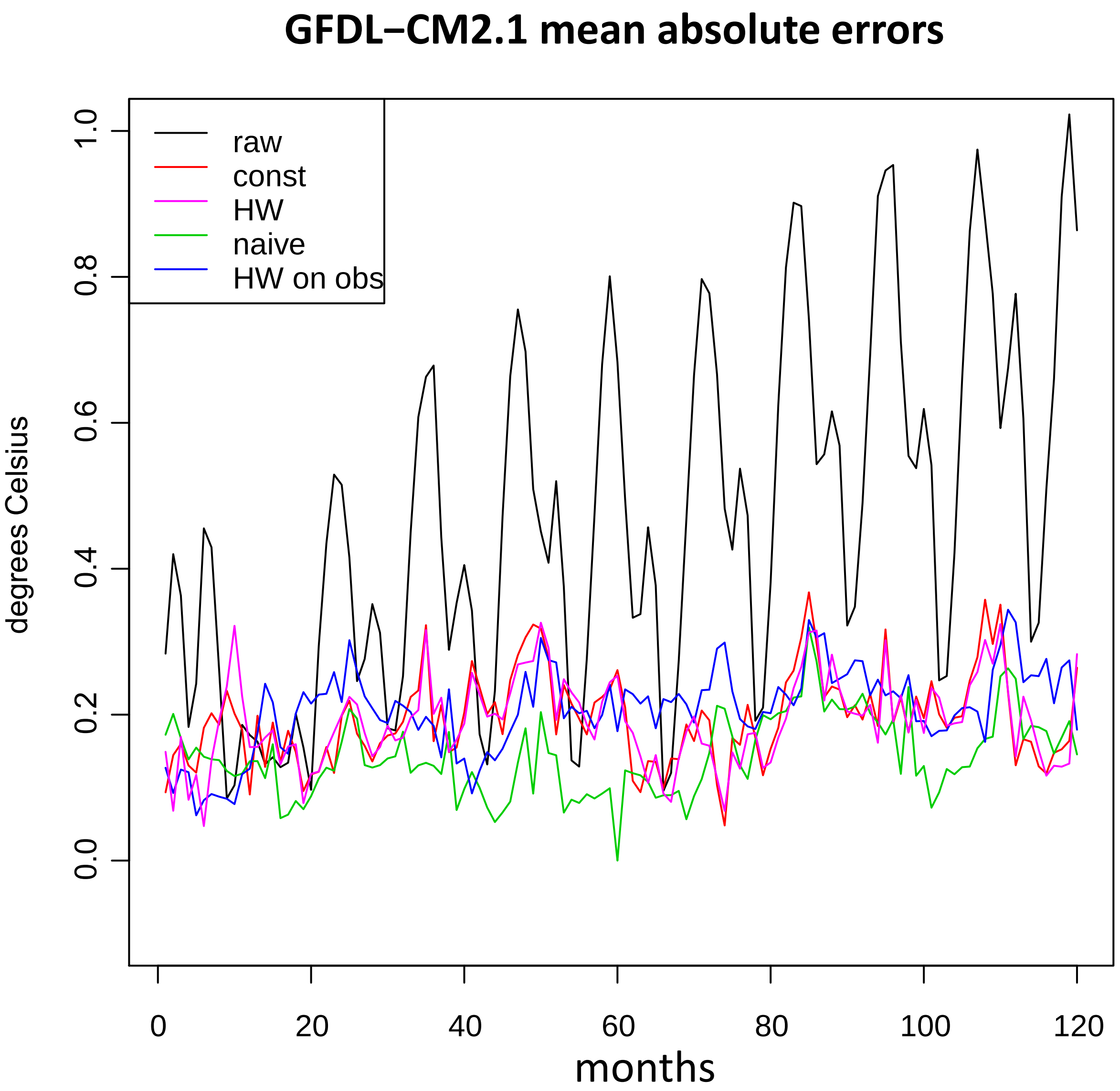
Bias correction methods

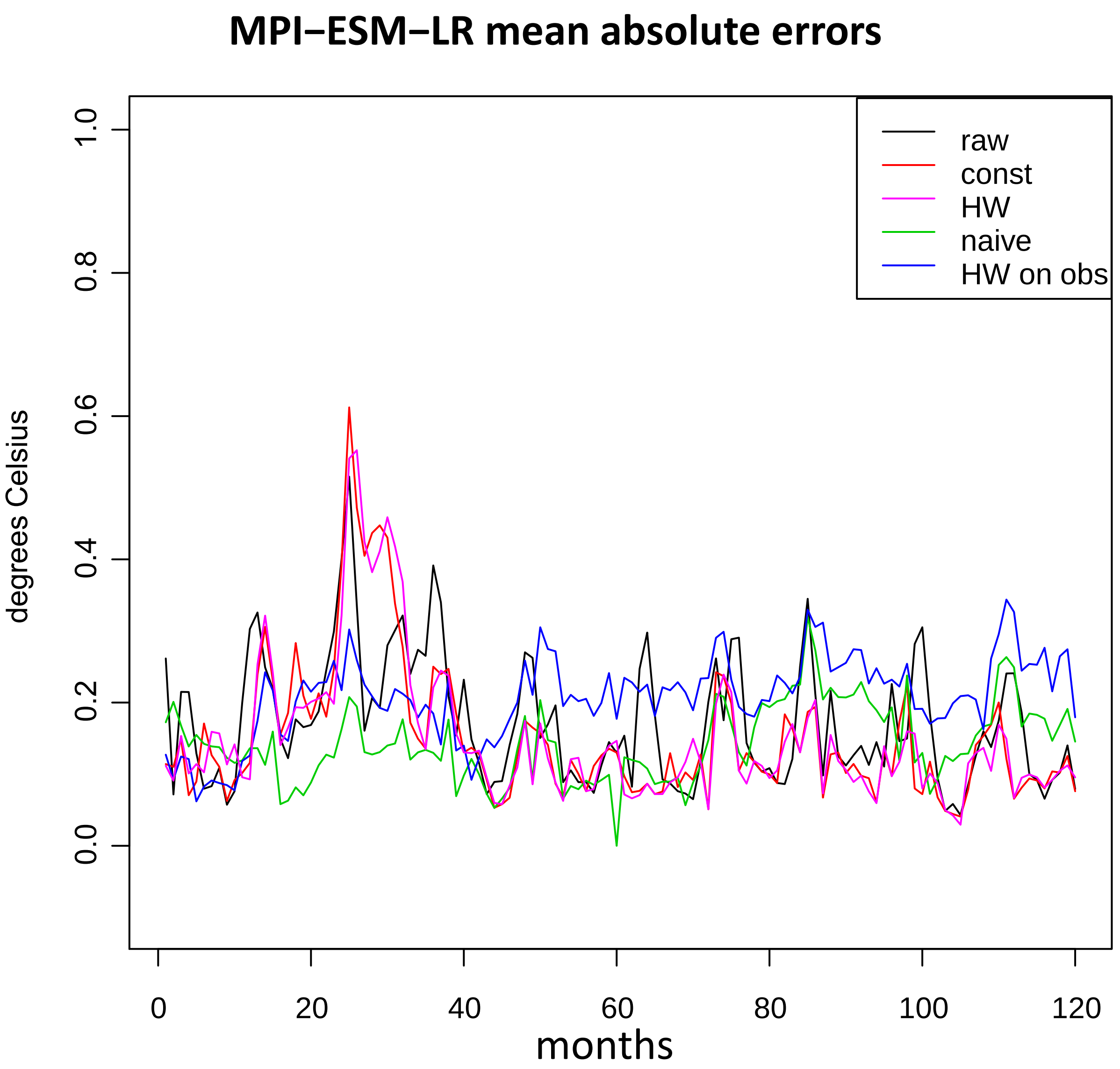
- Need to measure the difference between the observed series and the decadal forecast and use that measure to adjust (bias-correct) the raw decadal forecasts.
  - Mean adjustment (Garcia-Serrano and Doblas-Reyes, 2012)
    - For each forecast step, collect errors (= observed - raw forecast) from all available start dates and average them. Add the average error to future raw forecast to get bias-corrected forecast for that forecast step.
  - Exponentially smoothed adjustment
    - Fit exponential smoothing model to 120 months of a decadal forecast. Repeat as more start dates become available and average the fits. Use the average fitted values to get bias-corrected forecasts. (Uses ets module in R “forecast” package to find best exponential smoothing method fit.)

Comparisons: mean absolute values

- Calculate errors from bias-corrected forecasts, 1976-2001 start dates
  - For mean (constant) adjustment and for exponential smoothing adjustment
- For each forecast step, 1 to 120 months, obtain absolute error
- Average over all available start dates
- For naïve forecast, last available observed anomaly is the forecast
- For exponential smoothing on observations, fit to 120 months of observations then use result to forecast 120 months ahead and repeat through all available start times.

Geophysical Fluid Dynamics Laboratory (full-field initialization) and Max Planck Institute, Earth System Model (anomaly initialization) runs

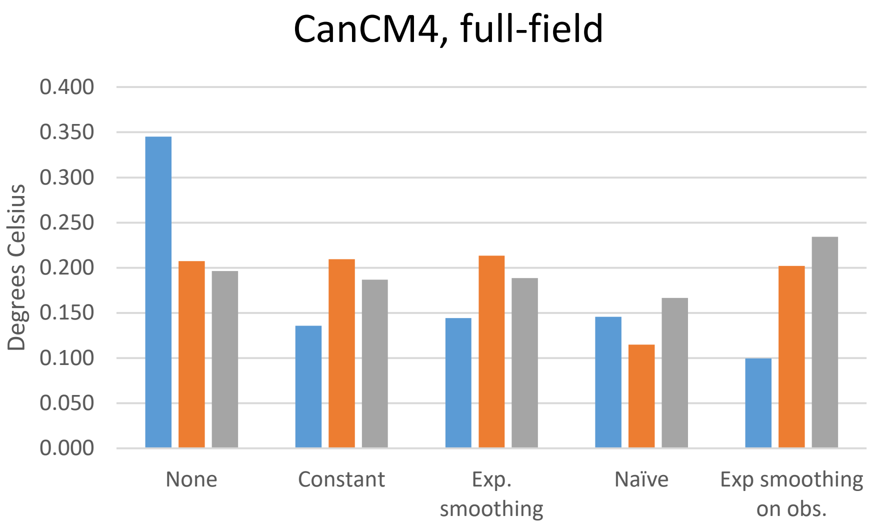


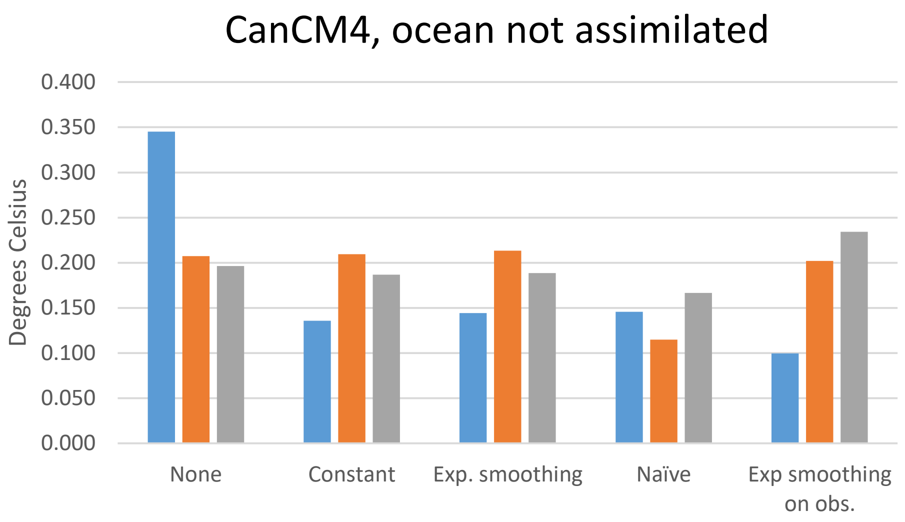


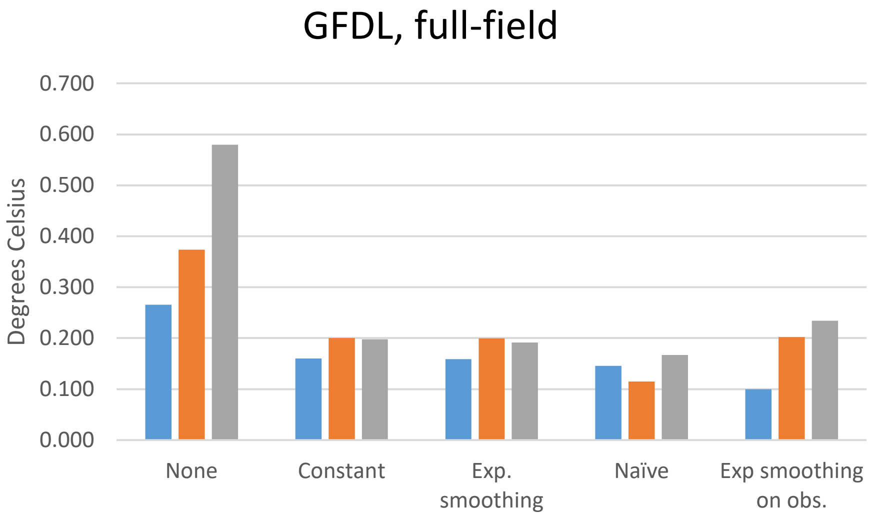
Number of times in 120-month forecast horizon the method has lowest mean absolute error

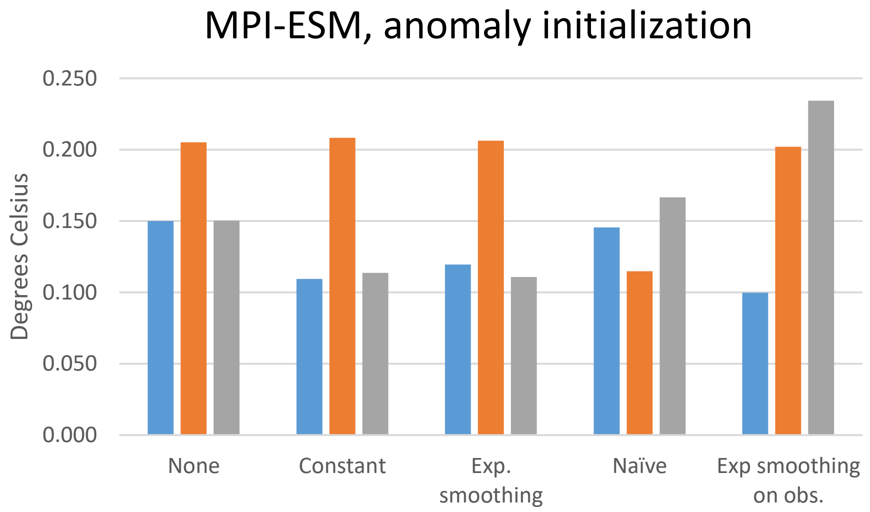
		Forecast method				
		GCMs			Statistical model	
Model	Initialization	raw	constant	exp smoothing	naïve	exp smoothing
CanCM4i1	Full field	24	14	12	59	11
CanCM4i2	Ocean unassimilated	18	25	20	46	11
GFDL	Full field	1	19	18	72	10
MPI-ESM	Anomaly	21	30	37	39	4
HadCM3i2	Anomaly	0	35	47	41	4
HadCM3i3	Full field	0	41	43	39	3

Mean absolute errors by years and groups of years:



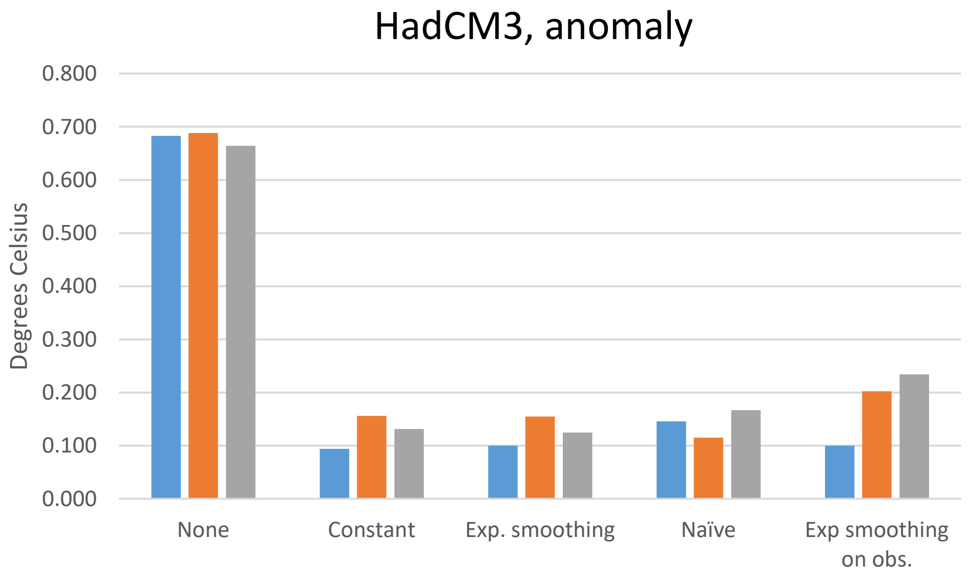


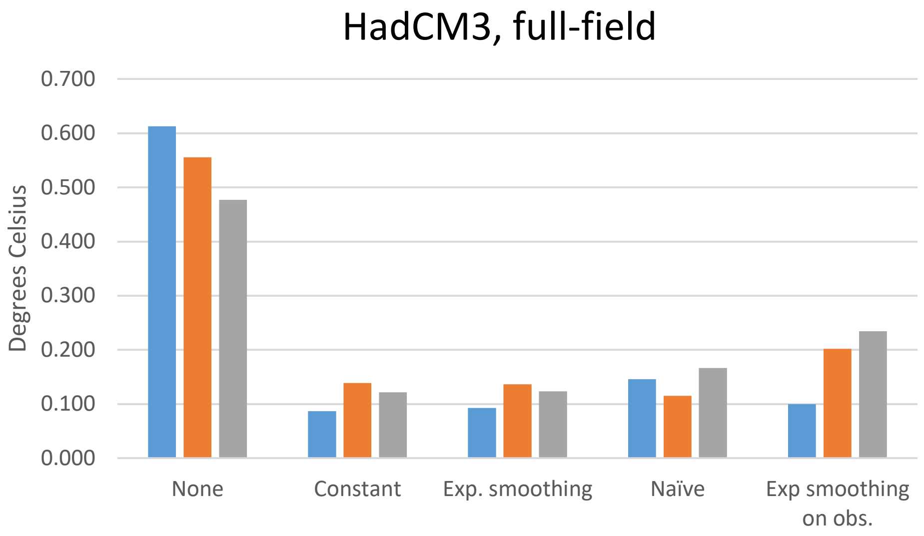




Mean absolute errors by years and groups of years:

Year ■ 1 ■ 2-5 ■ 6-10





Mean absolute errors of different bias corrections on different models

Year 1	Operation	CanCMi1 Full-field	CanCMi2 Not ocean	GFDL Full-field	HadCMi2 Anomaly	HadCMi3 Full-field	MPI-ESM Anomaly
None		0.345	0.383	0.265	0.683	0.613	0.150
Constant	De-bias	0.136	0.148	0.160	0.094	0.087	0.109
Exp. smoothing	De-bias	0.144	0.149	0.159	0.100	0.092	0.119
Naïve no change	On obs.	0.146	0.146	0.146	0.146	0.146	0.146
Exp. smoothing	On obs.	0.100	0.100	0.100	0.100	0.100	0.100
Years 2-5							
None		0.208	0.218	0.374	0.688	0.556	0.205
Constant	De-bias	0.210	0.180	0.200	0.156	0.139	0.208
Exp. smoothing	De-bias	0.213	0.180	0.199	0.155	0.136	0.206
Naïve no change	On obs.	0.115	0.115	0.115	0.115	0.115	0.115
Exp. smoothing	On obs.	0.202	0.202	0.202	0.202	0.202	0.202
Years 6-10							
None		0.196	0.222	0.580	0.664	0.477	0.150
Constant	De-bias	0.187	0.159	0.198	0.131	0.122	0.114
Exp. smoothing	De-bias	0.189	0.158	0.191	0.124	0.123	0.111
Naïve no change	On obs.	0.167	0.167	0.167	0.167	0.167	0.167
Exp. smoothing	On obs.	0.234	0.234	0.234	0.234	0.234	0.234

Conclusions

- Decadal forecasts from climate models are biased to varying degrees
- All appear to show greater amplitude than observation data
- Both methods of debiasing generally reduce error at all forecast steps
- Exponential smoothing method of debiasing no better than standard (separately calculated constant adjustment at each forecast step)
- Debiased GCM forecasts add nothing to observation-based forecasts **except at longer horizons**
  - **some GCMs are apparently better than others**
- Naïve-no-change forecast on observational data is a strong performer
- Errors on raw forecast and errors after de-biasing do not appear related (compare HadCM3 with MPI-ESM)

Other issues/ further research

- Combining GCMs
- Encompassing
  - Do GCMs add value to RW
- Conflict with other studies