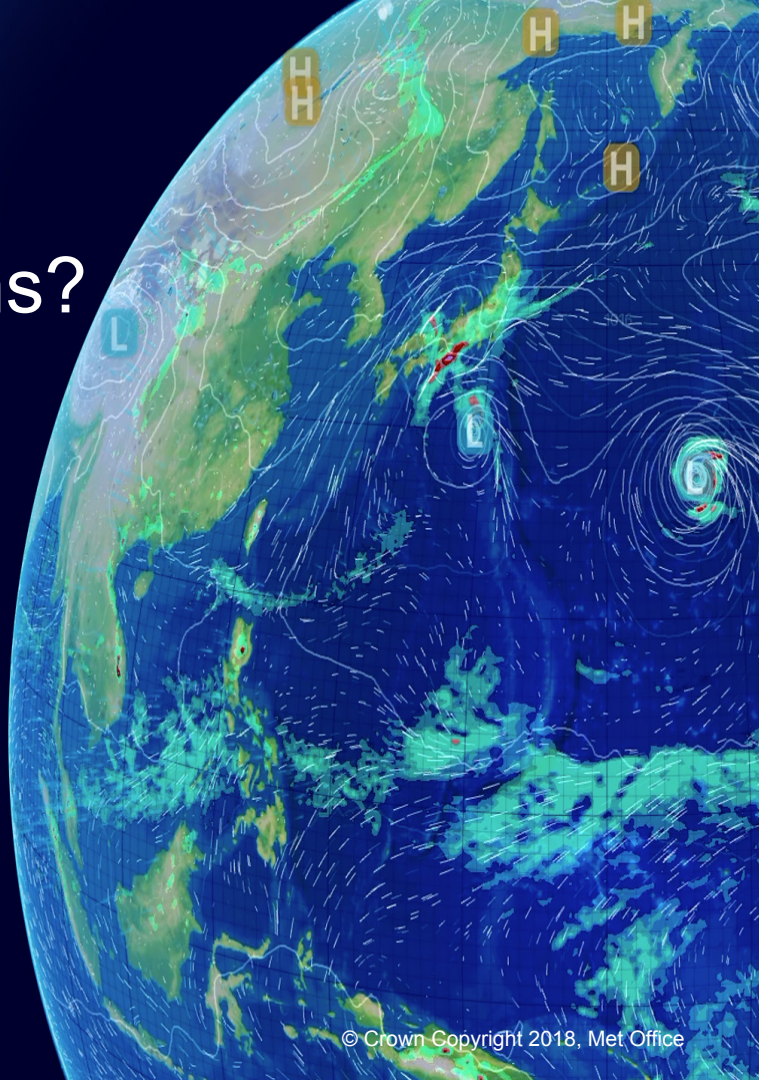
Met Office
Hadley Centre

How skilful are decadal predictions?

Doug Smith, Rosie Eade, Adam Scaife, Louis-Philippe Caron, Tom Delworth, Francisco J. Doblas-Reyes, Nick J. Dunstone, Leon Hermanson, Slava Kharin, Masahide Kimoto, Bill Merryfield, Takashi Mochizuki, Wolfgang Müller, Holger Pohlmann Xiaosong Yang and Stephen Yeager



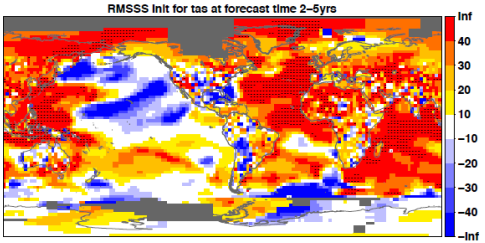
CMIP5 multi-model skill: years 2-5

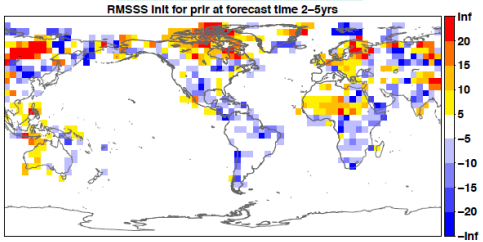
Temperature

Precipitation

Total skill (RMSSS)

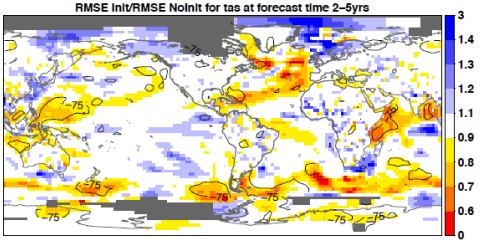
1 – RMSE / RMSE clim

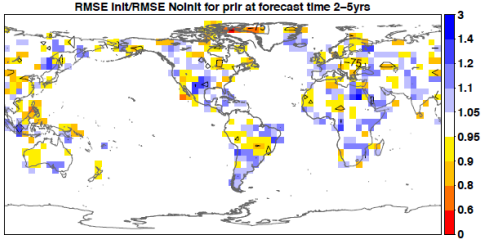




Impact of initialisation

RMSE init / RMSE NoInit





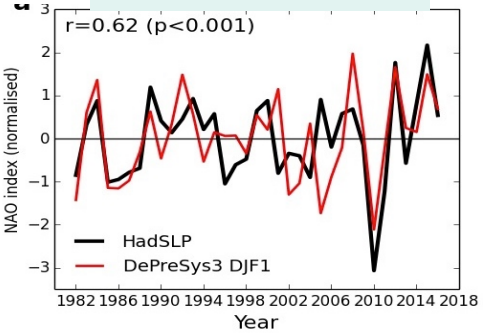
- High skill for temperature, limited skill for rainfall
- Improvement from initialisation mainly in North Atlantic, little impact over land

- Reassess skill in light of signal to noise paradox
  - Large ensemble
  - Focus on anomaly correlation
- Propose a more powerful method to assess the impact of initialisation

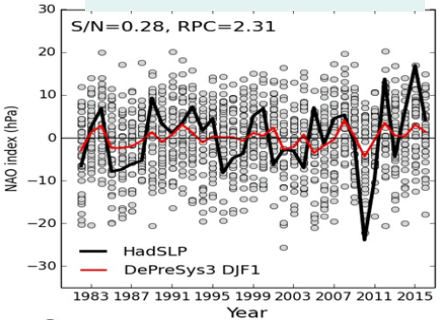
Kirtman et al 2013; Doblas-Reyes et al 2013

Signal to noise paradox

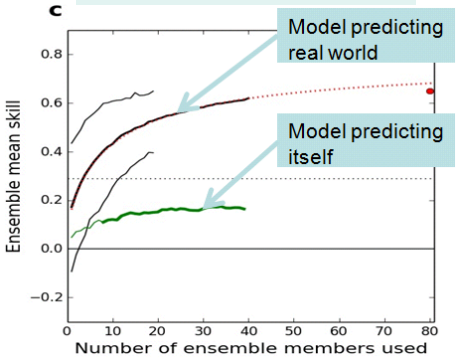
NAO: standardised



NAO: actual values



Impact of ensemble averaging



- Ratio of predictable components **RPC** = r(model~obs) / r(model~model)
- RPC > 1 implies:
  - Skilful forecasts possible using mean of **large ensemble** **(though paradoxically models cannot predict themselves!)**
  - but variability too small → **post processing** needed to adjust variance
  - skill measures of amplitude (e.g. RMSE, MSSS, probabilistic measures based on raw ensemble members) will **underestimate skill**
  - need **anomaly correlation** to assess available skill

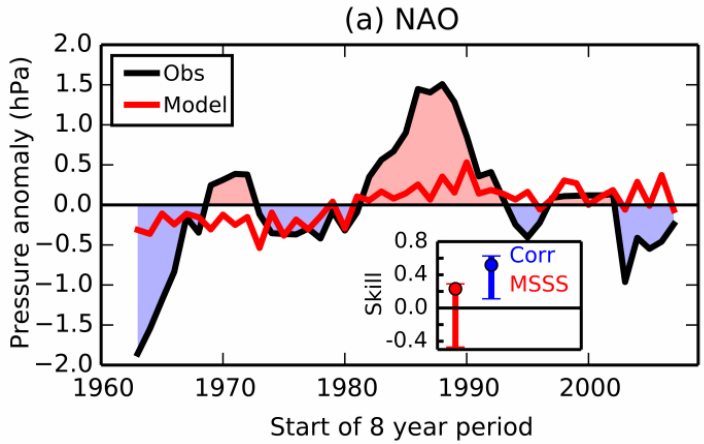
Scaife et al 2014, Eade et al 2014, Dunstone et al 2016, 2018, Baker et al 2018, Scaife and Smith 2018

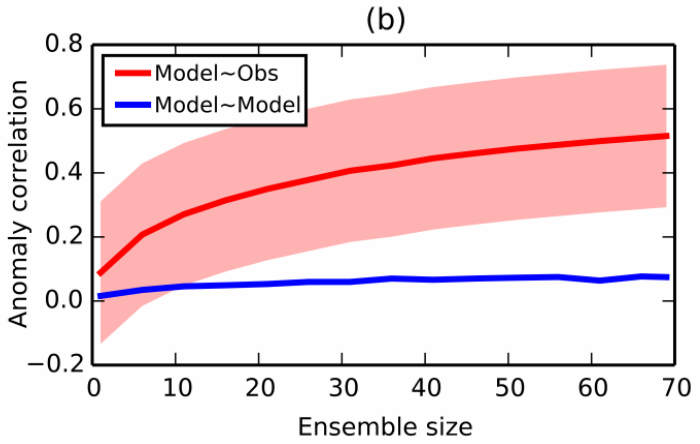
Assess skill of multi-model decadal forecasts

Forecast centre	Model	Initialised ensemble size	Uninitialized ensemble size
Met Office	HadCM3 (anomaly)	10	10
Met Office	HadCM3 (full field)	10	
CCCMA	CanCM4	10	10
GFDL	CM2	10	10
MIROC	MIROC5	6	3 (1 for precip/mslp)
MPI	MPI-ESM-LR	10	3
NCAR	CESM1.1	10	
BSC	EC-Earth	5	
	Total ensemble size:	**71**	**36**

- Hindcasts start dates every year from 1960
- CMIP5 plus new hindcasts from MPI, NCAR and EC-Earth
- All centres contribute to real-time multi-model exchange of decadal predictions

Skill: years 2-9: NAO (annual)

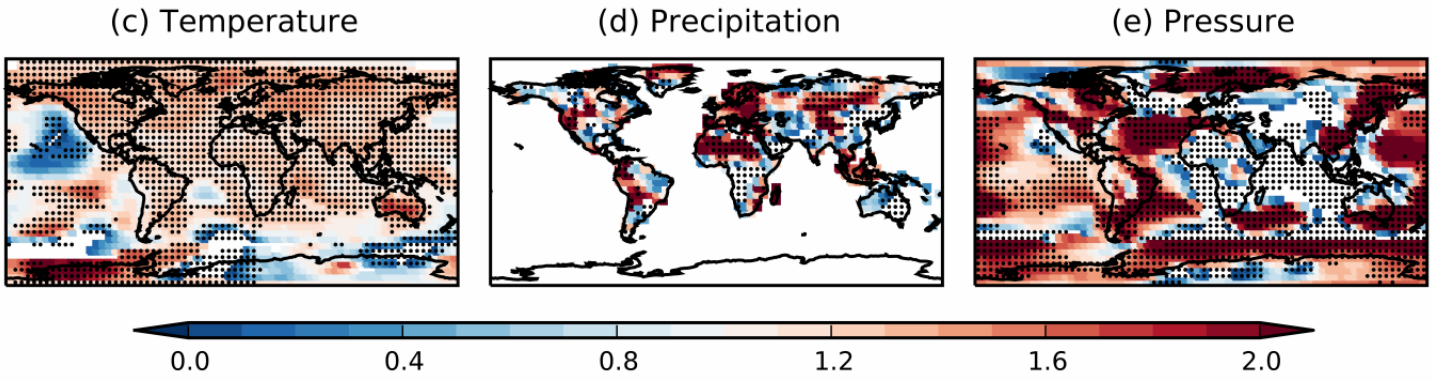




- Predicted signal has very small amplitude → **MSSS positive but not significant**
- BUT signal is somewhat similar to observations (increase from 1960s to 1990s, slight decrease thereafter)
- **Correlation is significant (r = 0.52, p = 0.027)**
- Skill is much higher with observations than with individual model members → **RPC > 6**

Smith et al, in prep

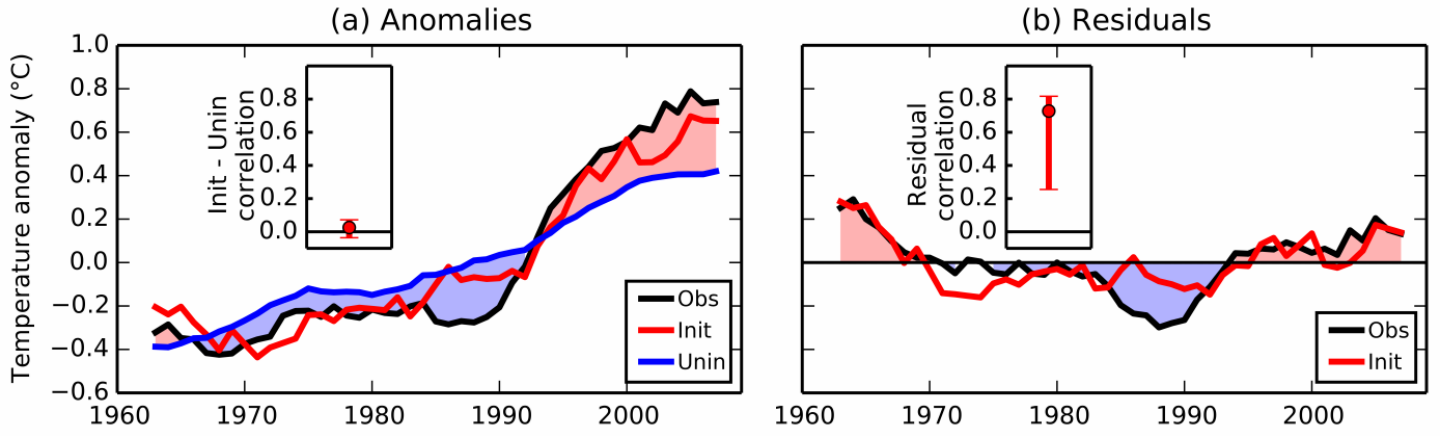
Ratio of predictable components (RPC): years 2-9



- RPC > 1 in many regions
- Especially for rainfall and pressure
- Signal to noise paradox is **widespread** on decadal timescales

Smith et al, in prep

Impact of initialisation: subpolar gyre temperature, years 2-9,JJA

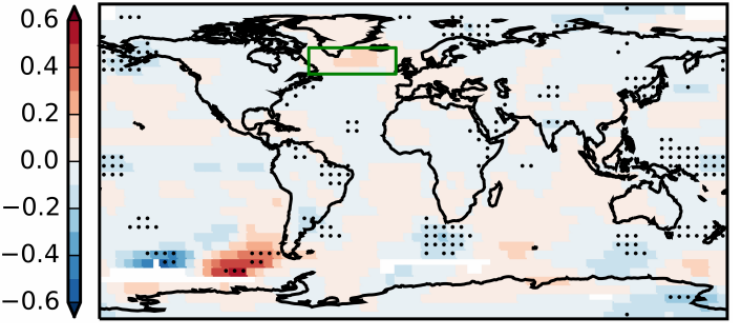


- Very high correlations for both initialised (Init r = 0.97) and uninitialized (Unin r = 0.94)
- Difference in correlations is not significant
- But residuals are highly and significantly correlated (r = 0.73, p = 0.007)
- **Initialised predictions capture some of the variability that is missing from uninitialized simulations**

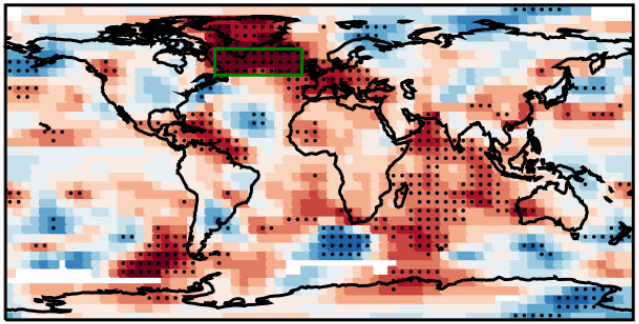
Smith et al, in prep

Impact of initialisation: temperature, years 2-9,JJA

Correlation difference



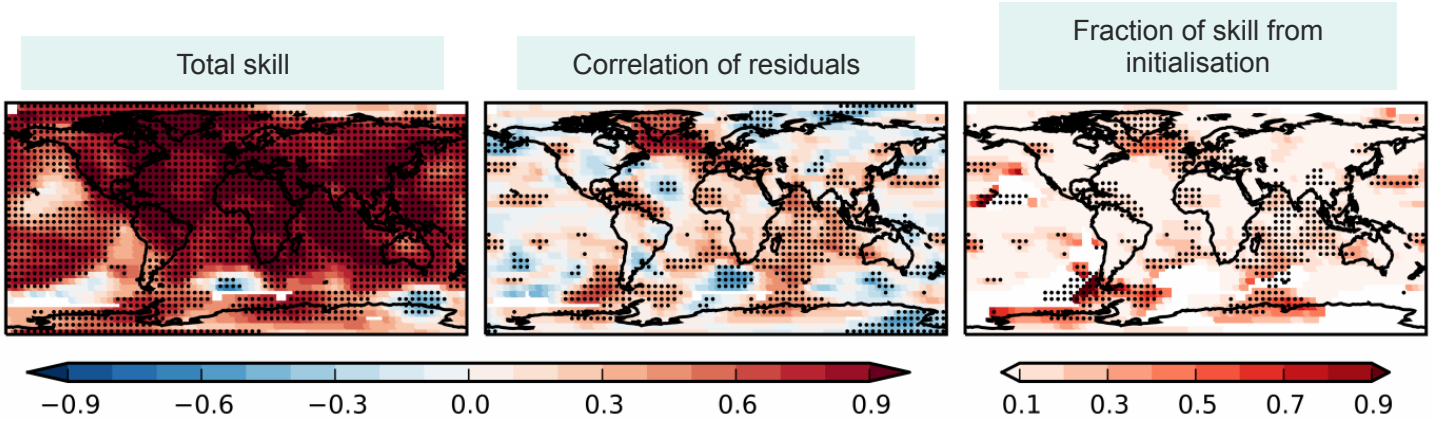
Correlation of residuals



- Improvement from initialisation is much clearer in correlation of residuals
- Impacts now seen over some land areas, including Europe

Smith et al, in prep

Impact of initialisation: temperature, years 2-9,JJA

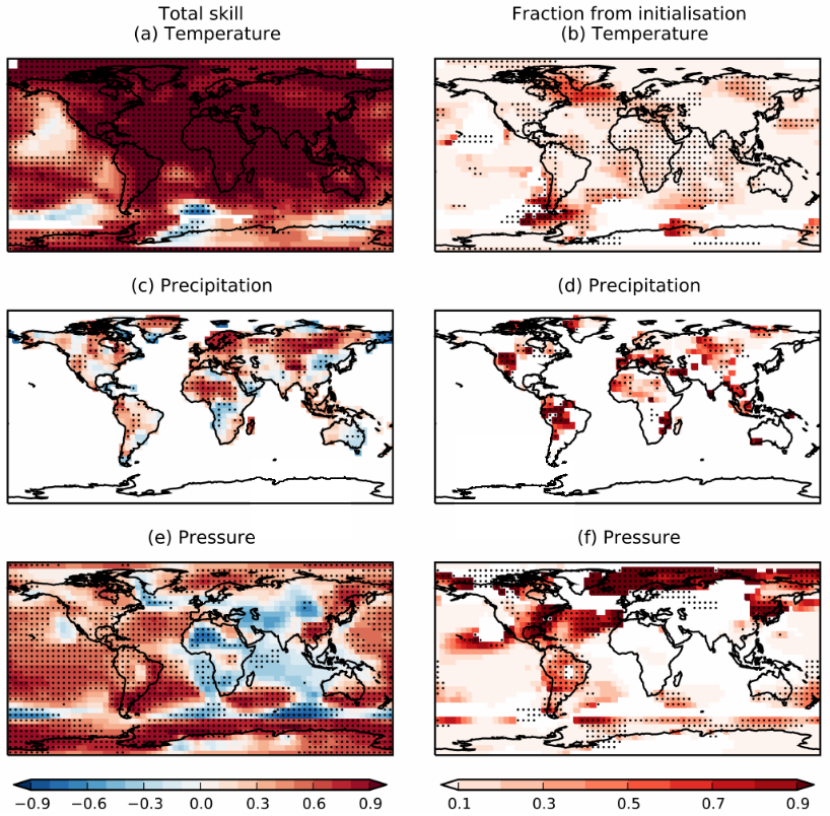


- Residuals may be correlated but unimportant if only a small part of the total variance
- Compute fraction of total skill coming from initialisation

Smith et al, in prep

Skill and impact of initialisation: years 2-9

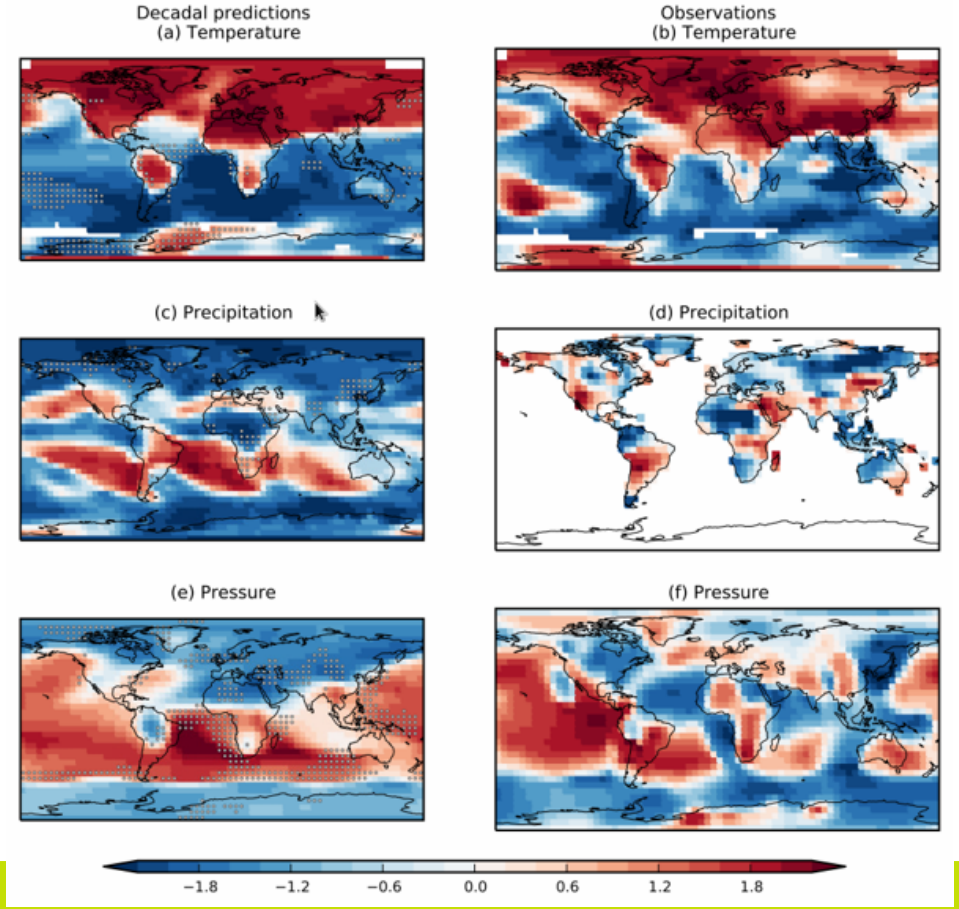
- High skill for temperature
- Significant skill for rainfall over land in many regions
- Significant skill for pressure (except Indian Ocean, central Asia, Africa)
- Significant improvements from initialisation
→ especially rainfall and pressure



Smith et al, in prep

Predicting regional patterns

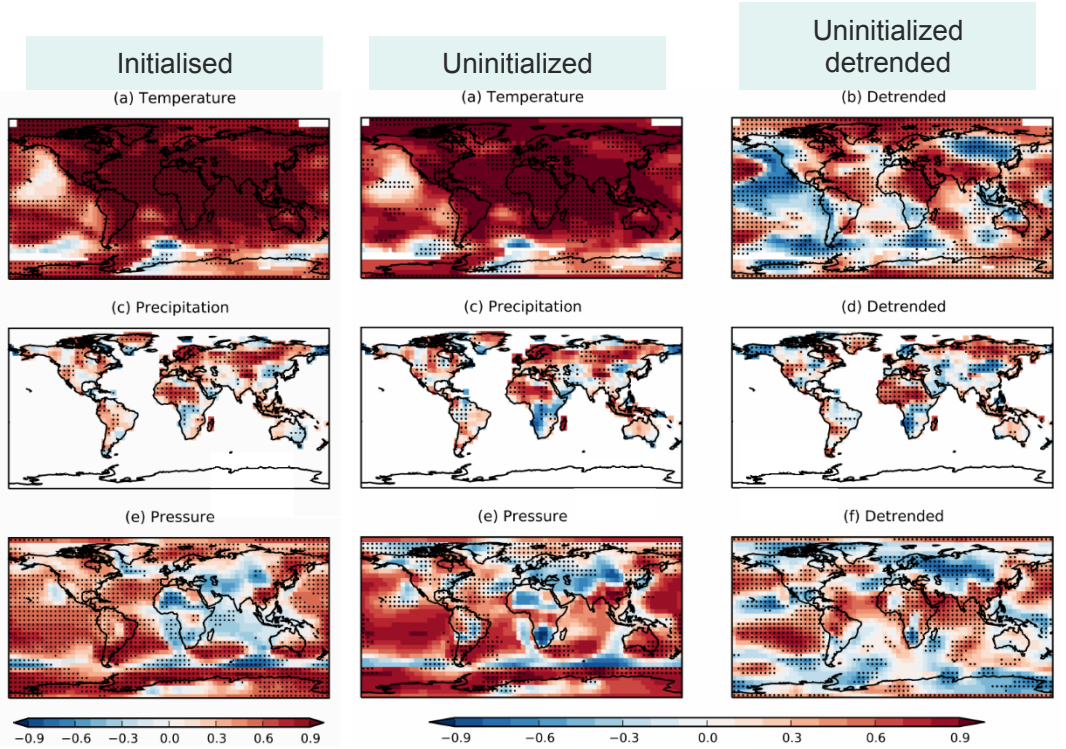
- Difference between1998-2014 and 1978-1994
- Both AMV and PDV changed sign
- Standardised anomalies
- Global average removed for temperature



Smith et al, in prep, following Robson et al papers

Central role of external forcing?

- Initialised and uninitialized patterns of skill are almost **identical**
- Improvements mainly where uninitialized already has some skill
- Initialisation mainly improving the **forced response** rather than predicting internal variability?
- Not just a trend – role for aerosols, solar, volcanoes…



Smith et al, in prep

Summary

- Signal to noise paradox
  - Need very **large ensemble** to extract predictable signal
  - Many measures underestimate skill → **anomaly correlation** needed
- Impact of initialisation
  - Assessing differences between (very high) correlations is not optimal
  - Assess **variability not captured by uninitialized simulations** → more powerful
- Decadal predictions are **skilful**
  - Temperature
  - Rainfall over land
  - Pressure
- Patterns of skill are very similar in uninitialized simulations
  - Initialisation is mainly **improving the response to external forcing?**
  - Detrended skill – role for **aerosols, solar, volcanoes**…

Assessing the impact of initialisation

- **Assessing differences between (very high) correlations is not optimal**
- **RMSE (and other measures) underestimate skill if the signal to noise ratio is too small**

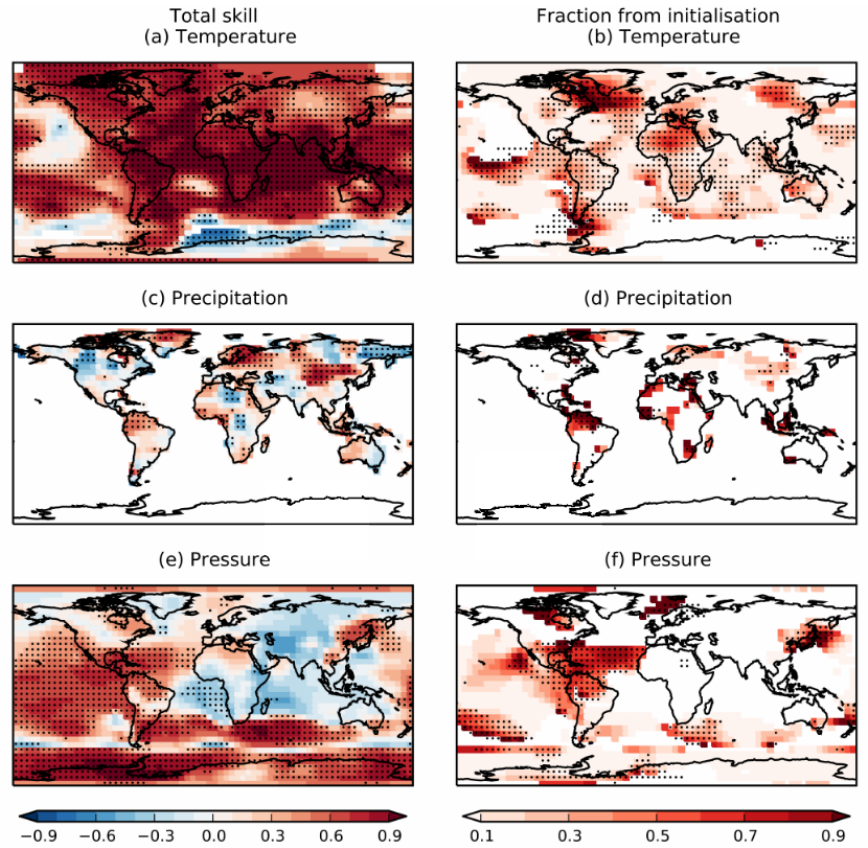
- Propose a new method: **does initialisation improve predictions of variability not captured by uninitialized simulations?**
  - Decompose forecast (f) and observed (o) time series

$$f = \hat{f} + f' \qquad o = \hat{o} + o'$$

  - Where $\hat{f}, \hat{o}$ are the components of *f* and *o* that are explained by linear regression of uninitialized simulations
  - The **residuals** f', o' are linearly independent of the uninitialized runs
  - Impact of initialisation may be assessed as correlation between residuals f', o'
  - This is likely to be larger than the differences between initialised and uninitialized correlations, increasing the "effect size" and enhancing the **power** of the test

Smith et al, in prep

Skill: years 2-9: DJF



Skill: years 2-9: JJA