# Do quantitative decadal forecasts from GCMs provide decision-relevant skill?

Emma Suckling and Leonard A. Smith

Centre for the Analysis of Time Series, London School of Economics

www2.lse.ac.uk/CATS

EGU 2012 Conference – Session CL3.3/NP5.4

25th April 2012

# Do decadal forecasts provide decision-relevant skill?

## Background

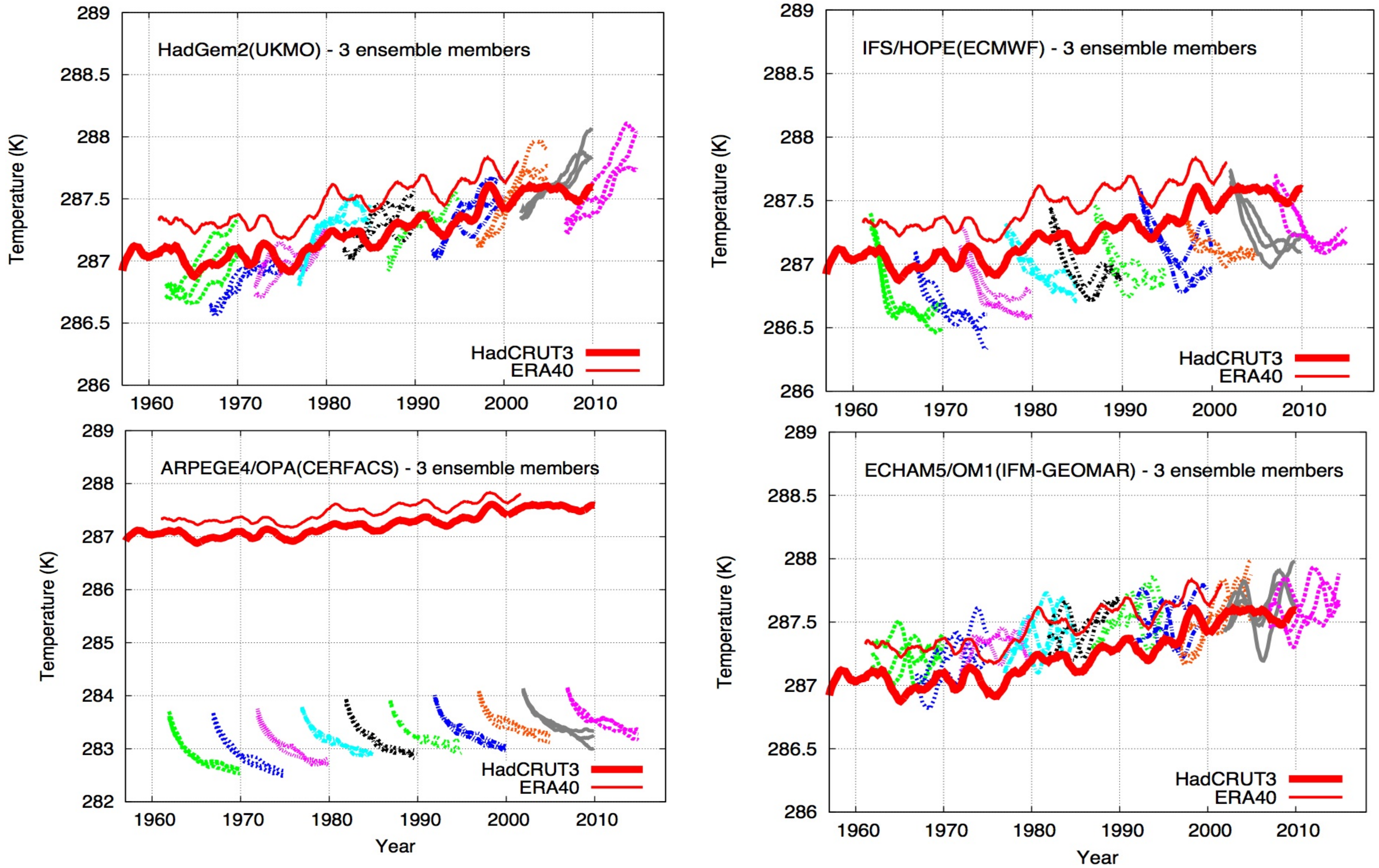
1. **Decadal predictions potentially provide useful information for decision makers**
    - **Adaptation to climate impacts**
    - **Flexible response planning**

2. **Decadal hindcast experiments and historical data archive allow standard forecast evaluation methods to be used**
    - **Skillful hindcasts do not guarantee predictability into the future**
    - **Evaluation of hindcast skill can inform future experimental design and impacts studies**
    - **Knowing about a lack of skill is also useful to decision-makers**

3. **Initialised hindcast experiments are a focus for IPCC AR5**
    - **What makes an appropriate benchmark to measure skill against for such forecasts?**

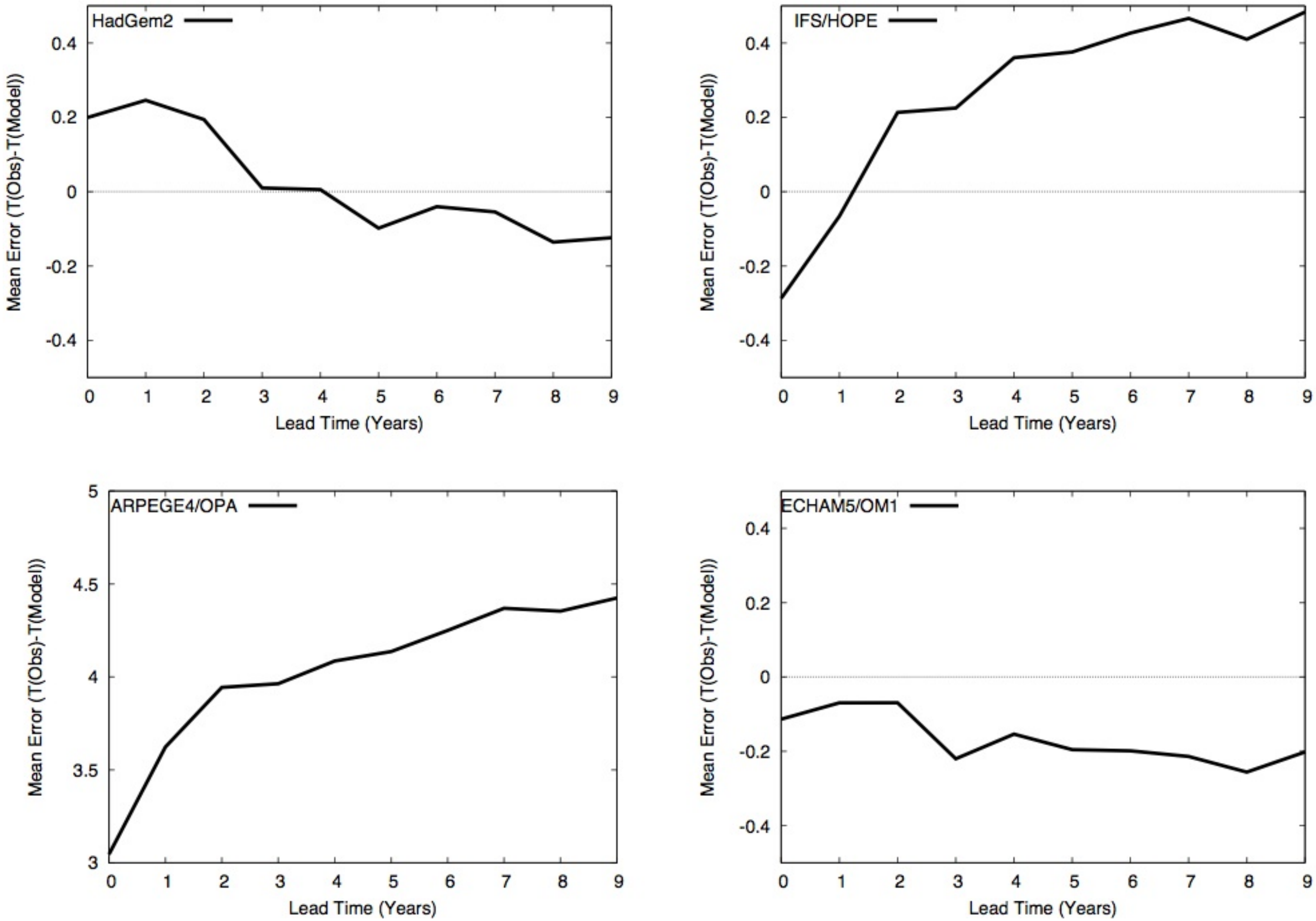
4. **Global mean temperature is the focus of this study**

# Decadal hindcast experiments - ENSEMBLES

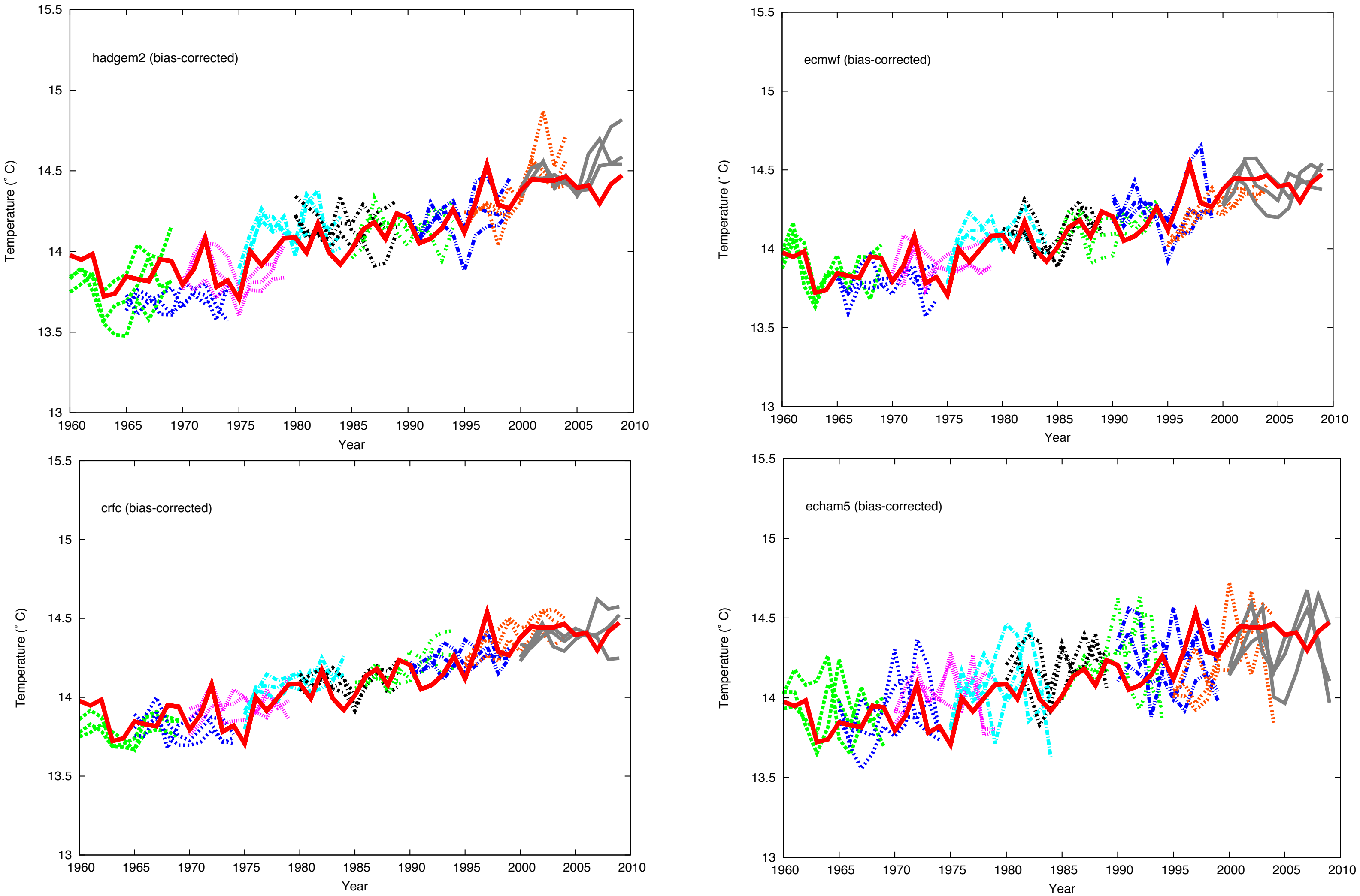



F. J. Doblas-Reyes *et al. Technical Memorandum* ECMWF, 621 (2010).

# Decadal hindcast experiments – model bias



# Decadal hindcast experiments after bias correction



# From ensembles to predictive distributions

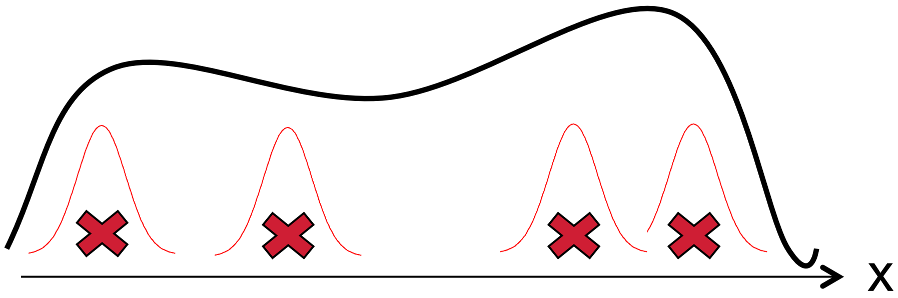
## Kernel Dressing

**The model-based component of the density, with *N* ensembles**

$$p(y:x,\sigma)=\frac{1}{N\sigma}\sum_{i=1}^{N}K\left(\frac{y-(x^{i}+\mu)}{\sigma}\right)$$

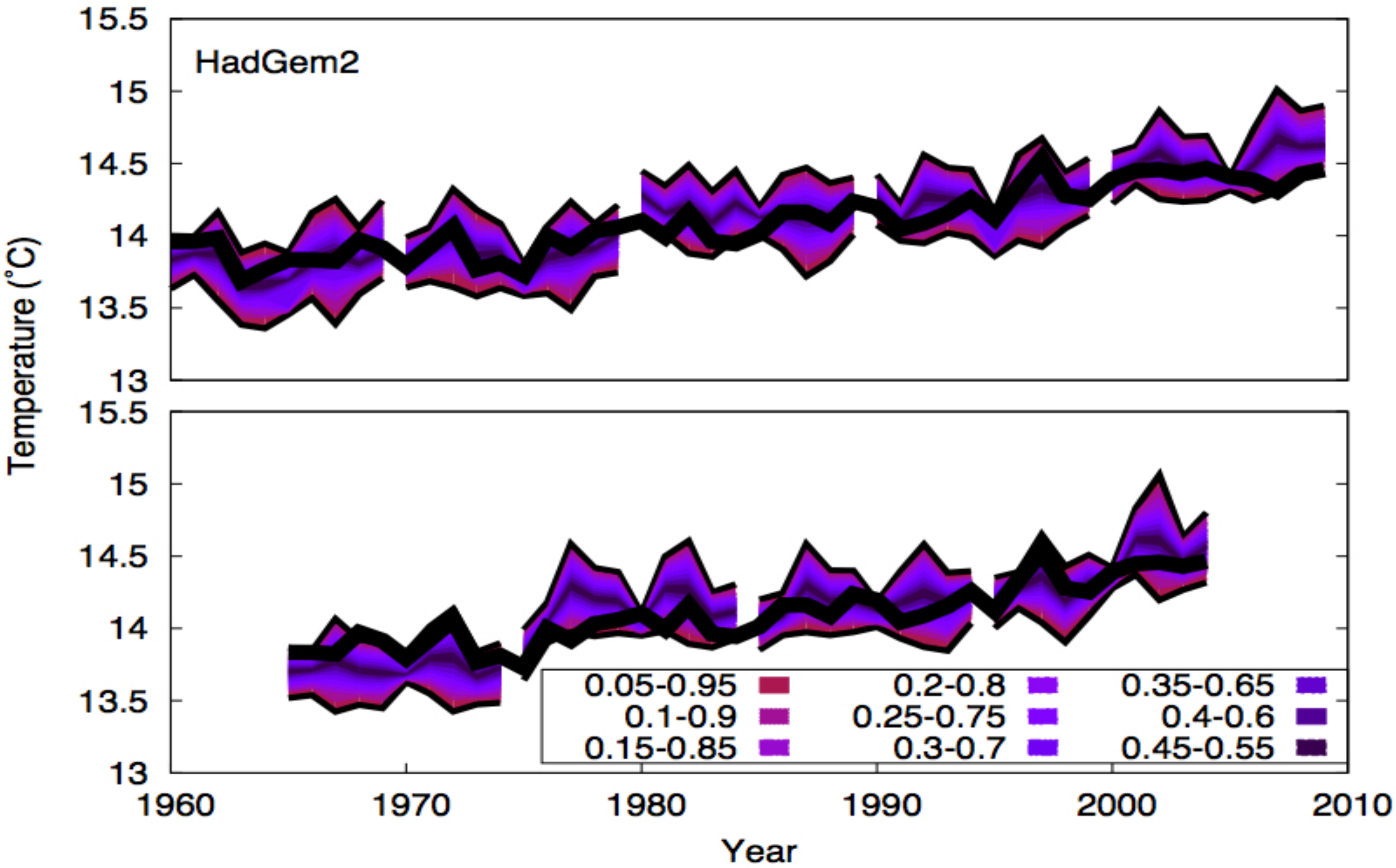
**Kernel, *K*, is a normalised Gaussian**

- **Parameters, μ (kernel mean plus bias correction) and σ (kernel spread), are varied to minimise the skill score (Ignorance, CRPS, quadratic)**
- **Cross-validation method important in evaluation**




**J. Bröcker and L. A. Smith, *Tellus A*, 60(4):663-678 (2007).**

# GCM decadal forecast distributions (GMT)




CATS CENTRE FOR THE ANALYSIS OF TIME SERIES

LSE THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

# Defining ‘zero skill’ with empirical models

## Dynamic climatology

1. **Evaluating forecast performance involves computing skill relative to a benchmark that defines ‘zero skill’**
   - Empirical models quantify our ability to predict without knowing the laws of physics
   - Climatology, persistence, statistical models serve as empirical benchmark models
   - But what makes an appropriate ‘zero skill’ model?

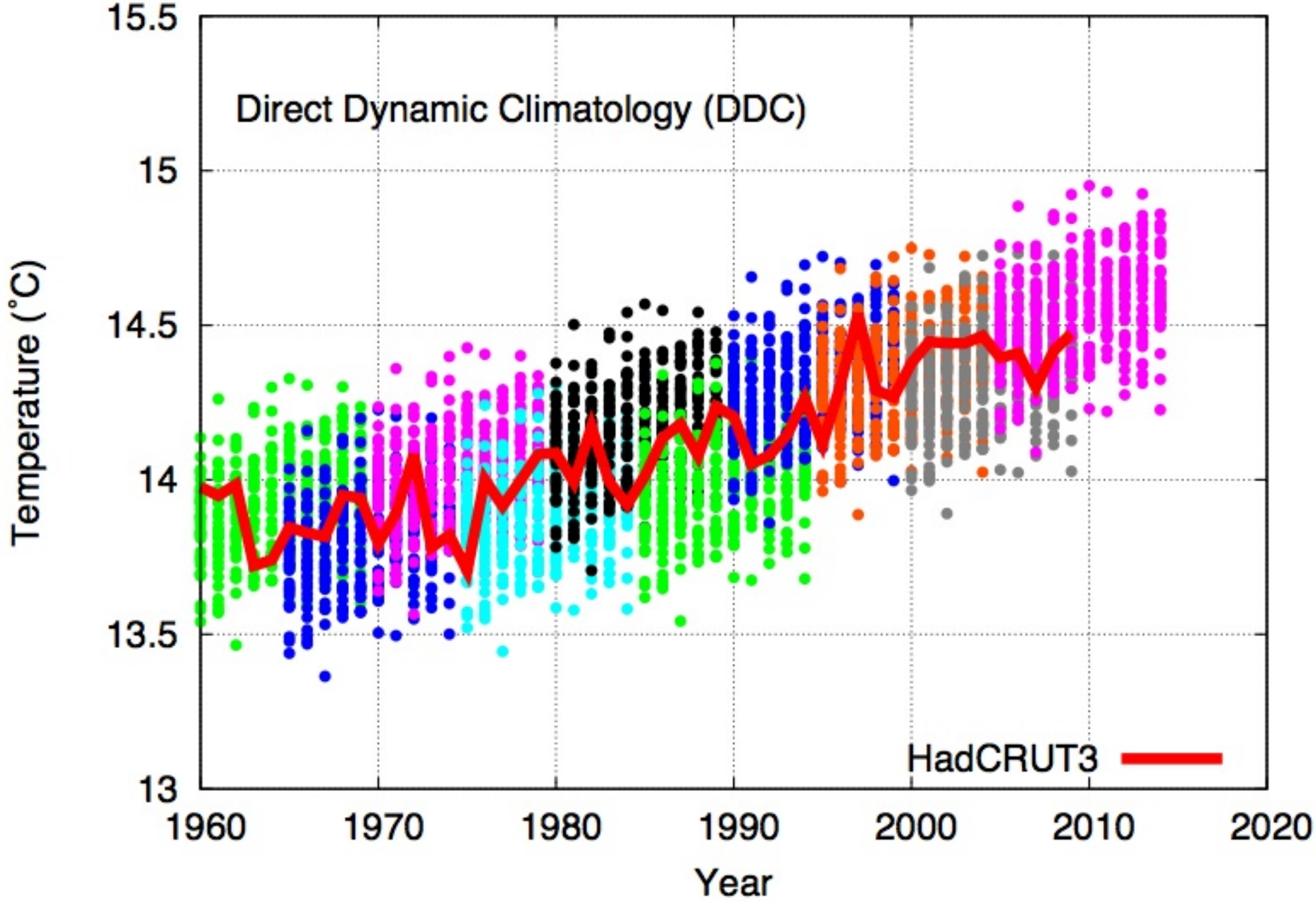
2. **Dynamic climatology (DC) is a more appropriate benchmark for near-term (initialised) climate forecasts**
   - A conditional climatology, initialised at launch and built from the historical archive
   - Ensemble members are composed of analogues of the current state, mapped forward in time
   - Analogues defined in terms of *l*th differences in observed record

$$f(l) = \sum_{i=1}^{N} O(t=0) + \Delta_l O(l_i)$$

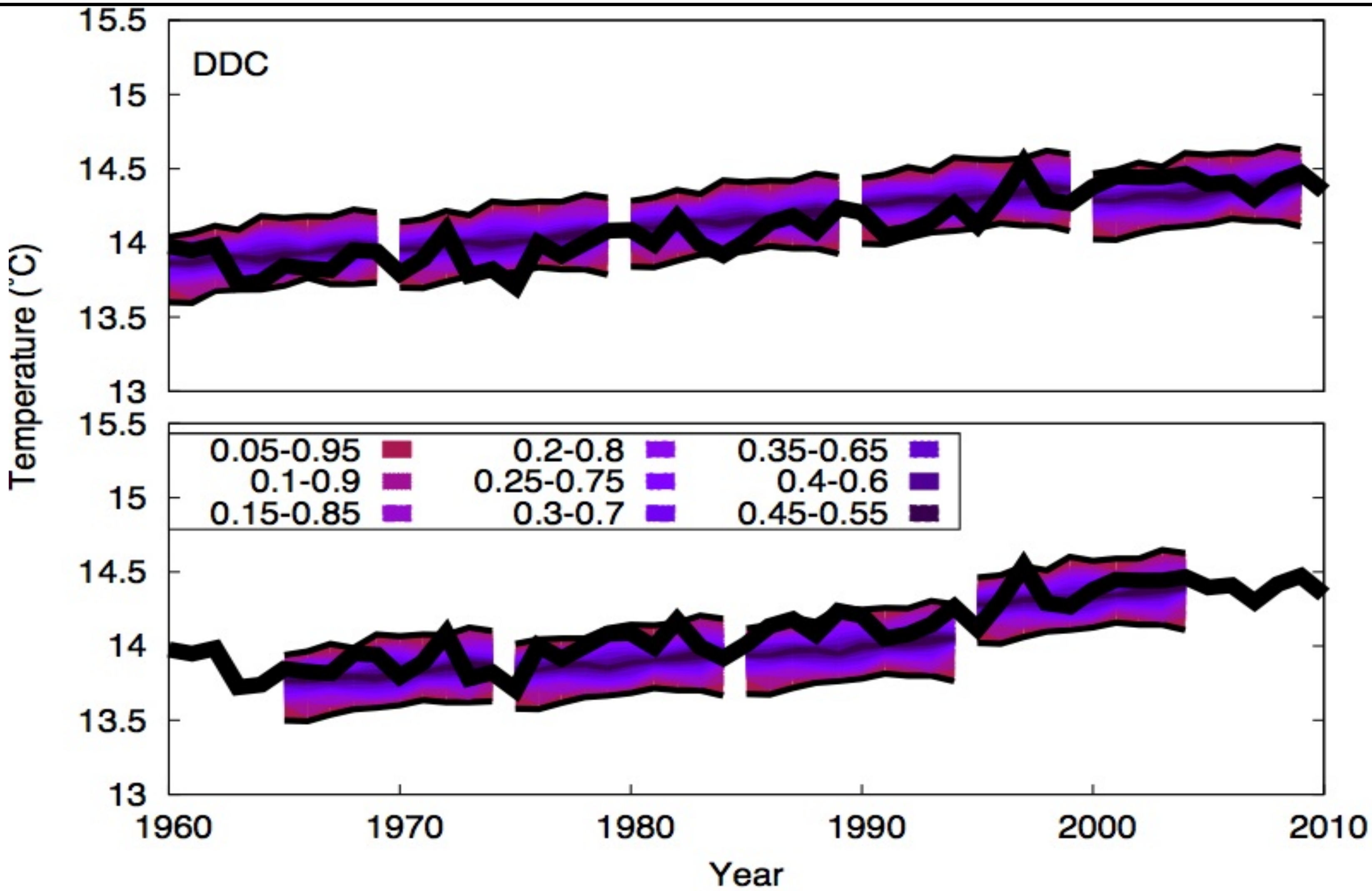
   - Choosing analogues:
     - Use full set of available analogues (dynamic climatology)
     - Define nearest neighbours and sample (random analogue prediction)

L. A. Smith, *Nonlinearity in Geophysics and Astrophysics*, CXXXIII:177-246 (1997).

# Dynamic climatology as a 'zero skill' benchmark

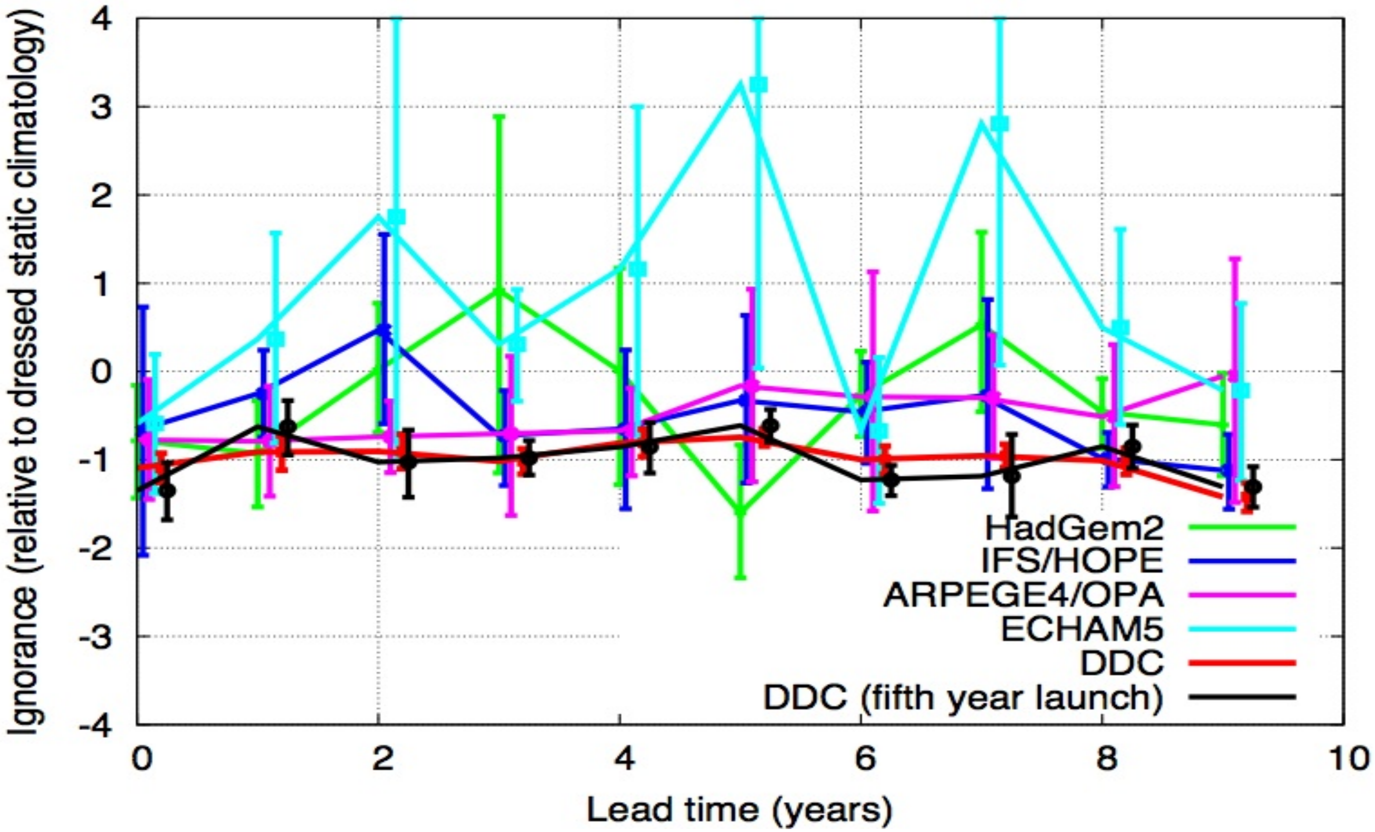


# Dynamic climatology forecast distribution (GMT)



# Forecast skill relative to static climatology

**Most models (including benchmark) show skill against static climatology**



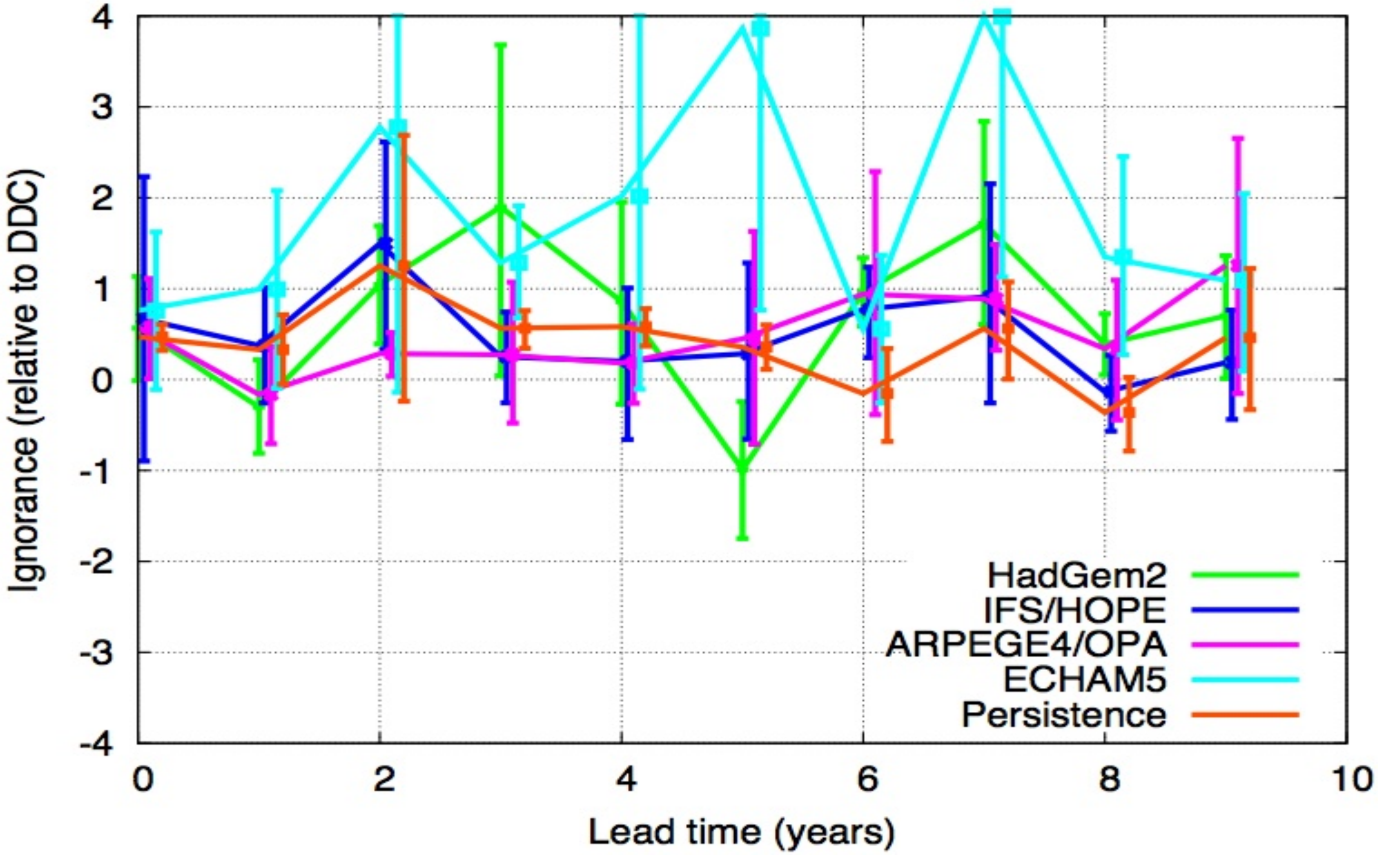

**Ignorance skill score:** $S_{rel}\left(p(y),Y\right)=\frac{1}{F}\sum_{i=1}^{F}-\log_2\left[\frac{p_i\left(Y_i\right)}{p_{ref}\left(Y_i\right)}\right]$

CATS CENTRE FOR THE ANALYSIS OF TIME SERIES

LSE THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

# Forecast skill relative to benchmark model

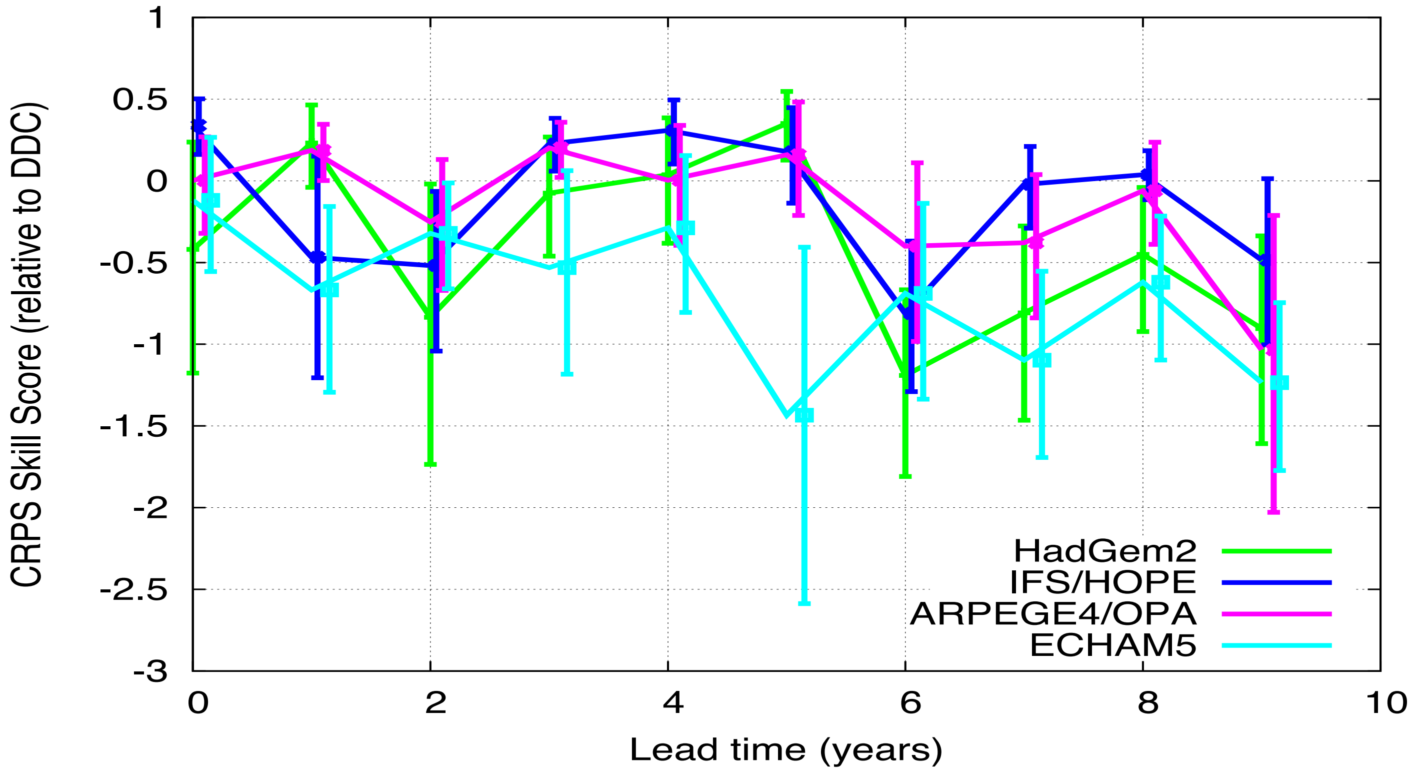
**No evidence of significant skill above the statistical benchmark**




**Similar results have been demonstrated over each of the Giorgi regions for surface air temperature at annual lead times**

# Evaluation of skill using CRP skill score

**Result is robust across skill scores**




**Continuous ranked probability score:** $CRPS = \int_{-\infty}^{y} G(x)^2 dx + \int_{y}^{\infty} \left[G(x) - 1\right]^2 dx$

**Skill score:** $SS = 1 - \dfrac{CRPS_{forecast}}{CRPS_{reference}}$

CATS CENTRE FOR THE ANALYSIS OF TIME SERIES

LSE THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

# Do decadal forecasts provide decision-relevant skill?

## Summary

1. **A robust and transparent forecast evaluation procedure is desirable**
   - **Important because of decision-relevance**
   - **Forecast skill is necessary, but not sufficient to establish value to a user**
   - **Evaluation and post-processing methods need to be made clear in advance**
   - **Appropriate measures of skill and value evaluated**

2. **A framework for in-sample skill evaluation of ensemble prediction systems is demonstrated**
   - **Approach addresses the value added compared to a 'zero skill' benchmark model**

3. **Significant skill is not found for any of the ENSEMBLES GCMs against a simple data-based model**
   - **Demonstrated at GMT and Giorgi region scale for ENSEMBLES models**
   - **Small sample size is part of the problem – results should inform future design**
   - **How do we use scientific insight given low model skill?**

**E. B. Suckling and L. A. Smith,** Do decadal predictions from GCMs yield decision-relevant skill? ***in preparation.***

# Do decadal forecasts provide decision-relevant skill?

## Methodology - Summary

1. **Bias correction**
   - Evaluation framework allows different bias correction approaches to be tested
   - Variational method seeks post-processing parameters that produce minimised score

2. **Kernel dressing**
   - Predictions for decision-making are generally given out as probabilistic forecasts
   - Evaluating post-processed forecasts provides a better basis for decision-makers

3. **Scoring rule**
   - Evaluating proper scoring rules provides most robust measures of skill
   - All scoring rules applied in this case led to the same conclusion

4. **Benchmark model**
   - Dynamic climatology is a more appropriate 'zero skill' model against initialised decadal hindcasts
   - Differences (changes) are chosen rather than absolute values and care is taken to not over fit or cheat by leaving out period being forecasted/evaluated
   - More sophisticated methods to choose analogues could be applied
   - Initialised DDC forecast for the next decade is based on differences over last decades

# Thank You!

## Contact Me

**Emma Suckling:** Centre for the Analysis of Time Series
London School of Economics

**e.suckling@lse.ac.uk**

## Related posters and talks

**CL3.3/NP5.4:** Insightful measures of predictive skill in seasonal forecasts
**H. L. Du, F. Niehörster & L. A. Smith** (Hall Z #83 – 17.30)

Skill scores for probabilistic climate and weather prediction
**T. Maynard, E. B. Suckling & L. A. Smith** (Hall Z #84 – 17.30)

**AS1.2:** Gradient descent assimilation for the point-vortex model
**E. B. Suckling and L. A. Smith** (Room 14 - 09.30)

***The support of NERC under the EQUIP project is gratefully acknowledged.***

CATS CENTRE FOR THE ANALYSIS OF TIME SERIES

EQUIP

LSE THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE