Sources of predictability and error in ECMWF long range forecasts

Tim Stockdale

European Centre for Medium-Range Weather Forecasts

Outline

- Overview of System 4
- Some recent research results

Seasonal prediction at ECMWF

- **Started in the 1990's**
- **Strategy: fully coupled global GCMs**
- **Real-time forecasts since early 1997**
  - Forecasts issued publicly from December 1997
- **Now using "System 4"**
  - Lifetime of systems has been about 5 years each

S1	S2	S3	S4
Dec 1997	Mar 2002	Mar 2007	Nov 2011

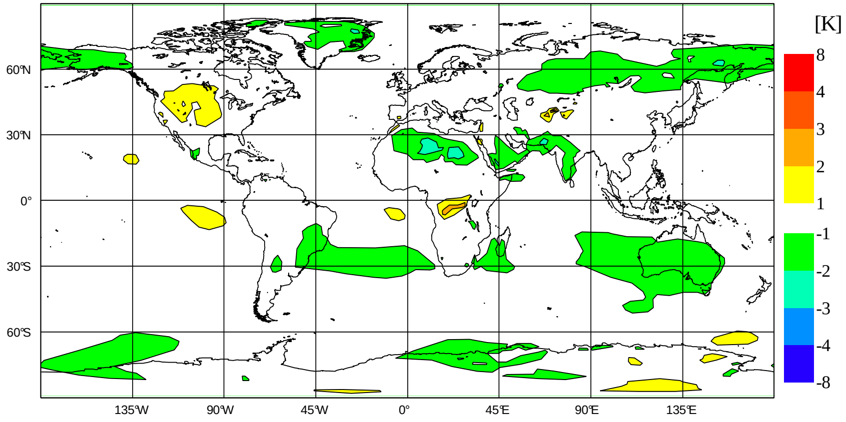
System 4 seasonal forecast model

- **IFS (atmosphere)**
  - T_L255L91 Cy36r4, 0.7 deg grid for physics (operational in Dec 2010)
  - Full stratosphere, enhanced stratospheric physics
  - Singular vectors from EPS system to perturb atmosphere initial conditions
  - Ocean currents coupled to atmosphere boundary layer calculations
- **NEMO (ocean)**
  - Global ocean model, 1x1 resolution, 0.3 meridional near equator
  - NEMOVAR (3D-Var) analyses, newly developed.
- **Coupling**
  - Fully coupled, no flux adjustments
  - Sea-ice based on sampling previous five years

Reduced mean state errors

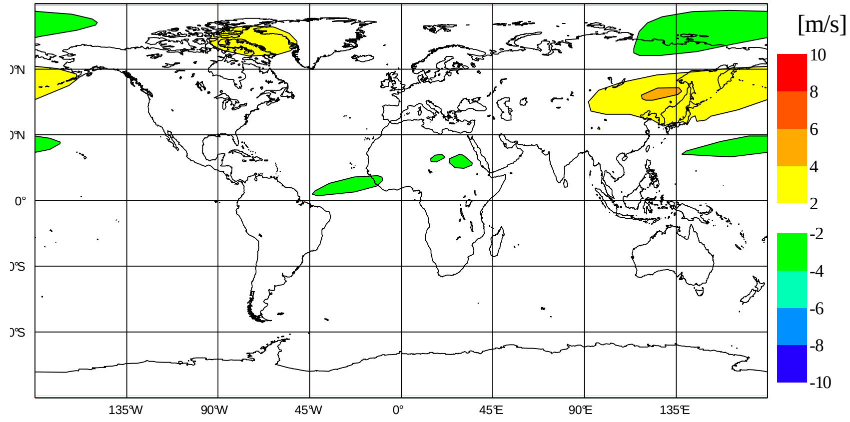
T850

850hPa temperature S4(15)-ERA Int 1991-2008 JJA
Global rms error: 0.663 NH:0.669 TR:0.662 SH:0.66



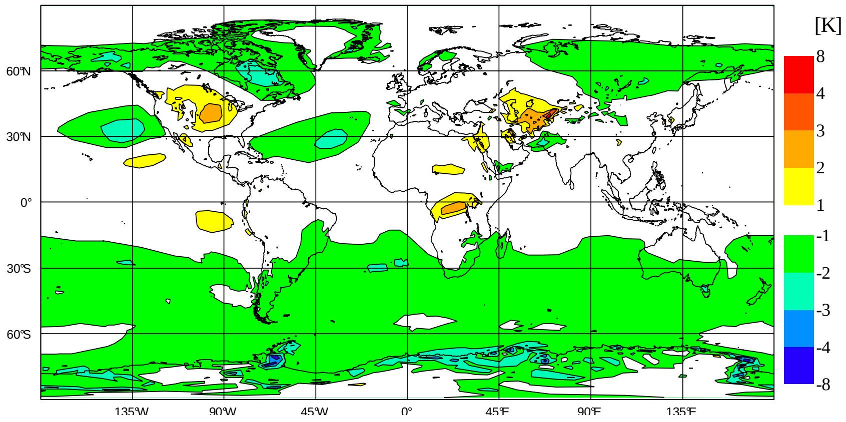
U50

50hPa zonal wind S4(15)-ERA Int 1991-2008 DJF
Global rms error: 1 NH:1.43 TR:0.853 SH:0.72

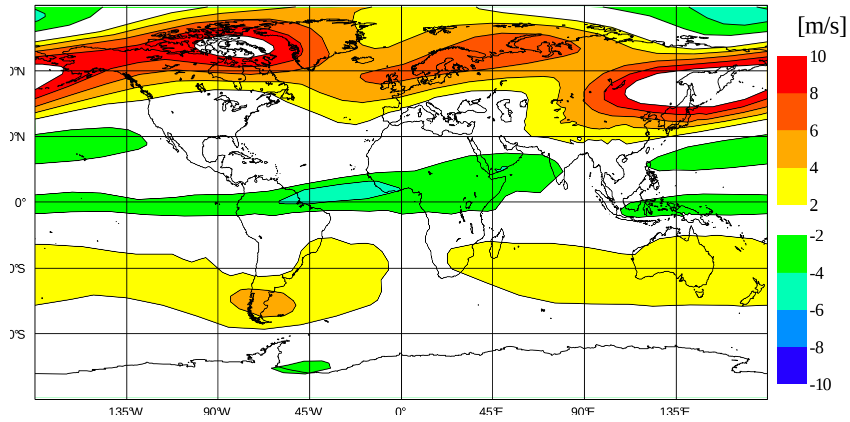


S4

850hPa temperature S3(11)-ERA Int 1991-2008 JJA
Global rms error: 1.07 NH:1.06 TR:0.798 SH:1.48



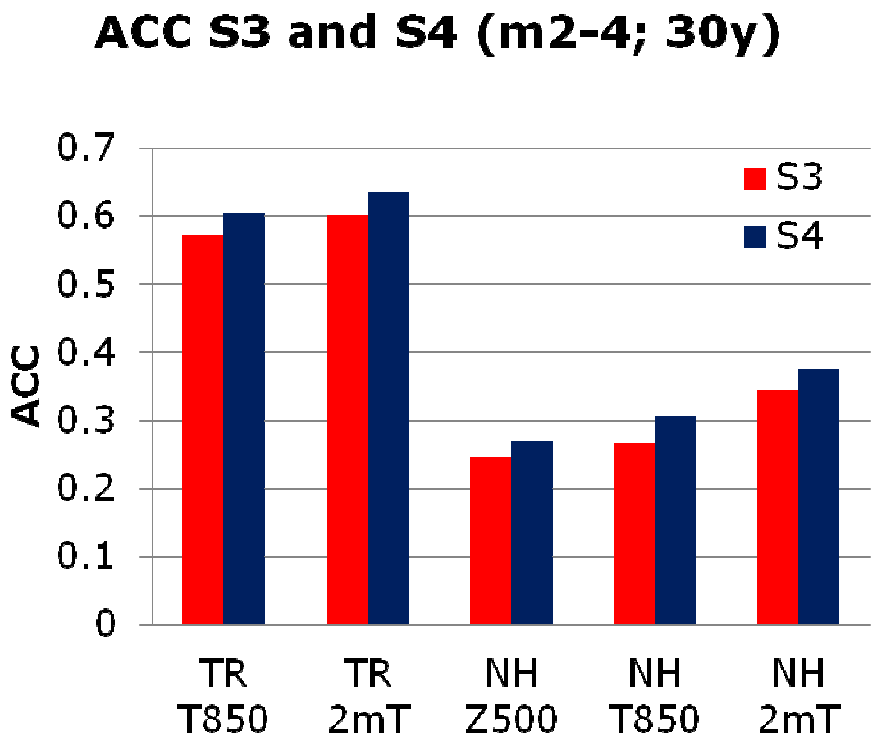
50hPa zonal wind S3(11)-ERA Int 1991-2008 DJF
Global rms error: 3.26 NH:5.53 TR:2.02 SH:2.03



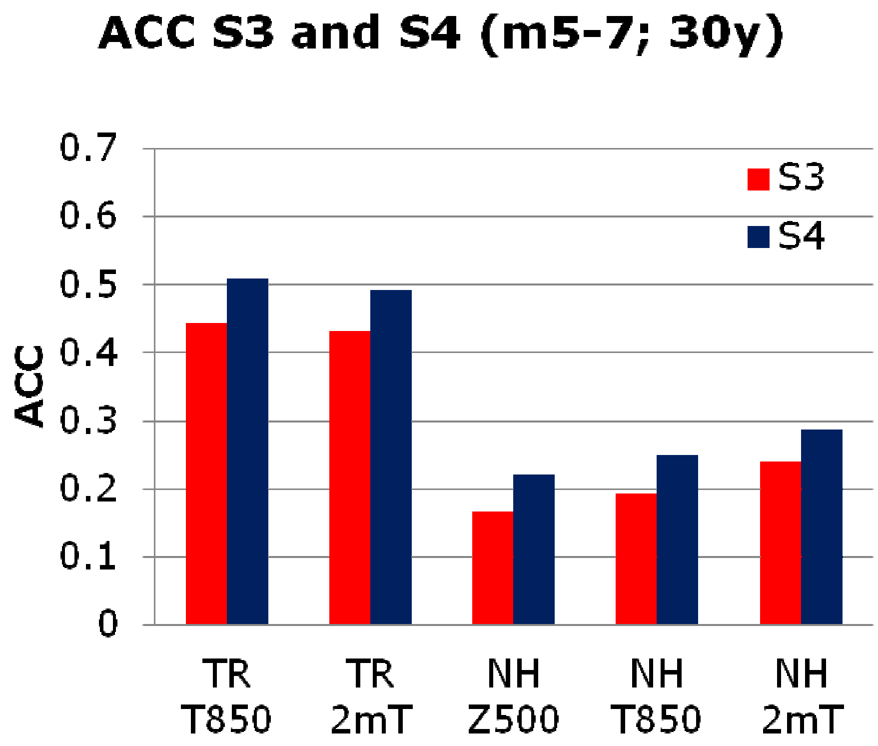
S3

Tropospheric scores

Spatially averaged grid-point temporal ACC



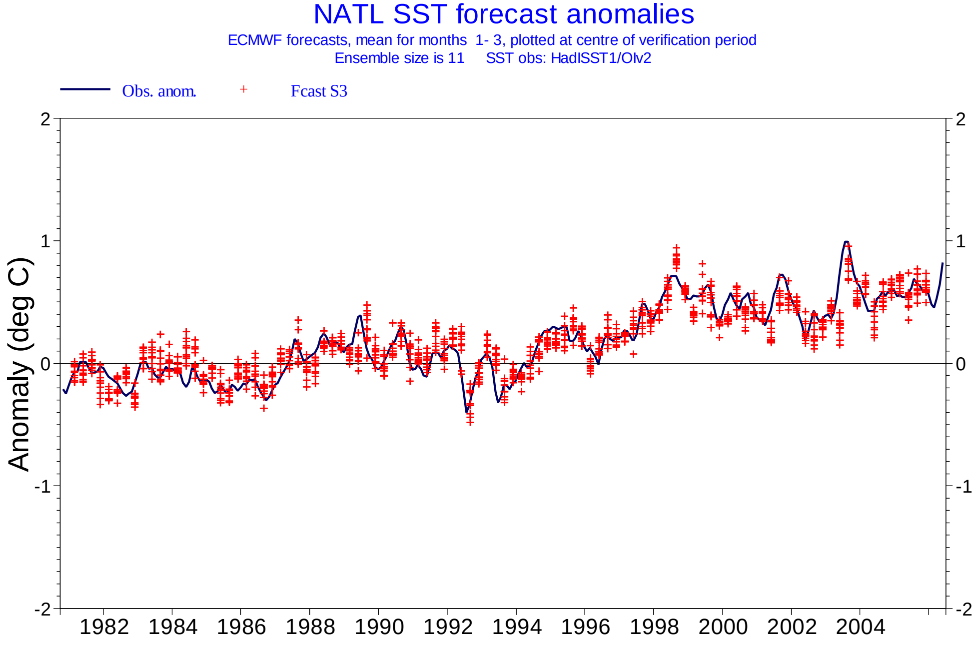
One month lead

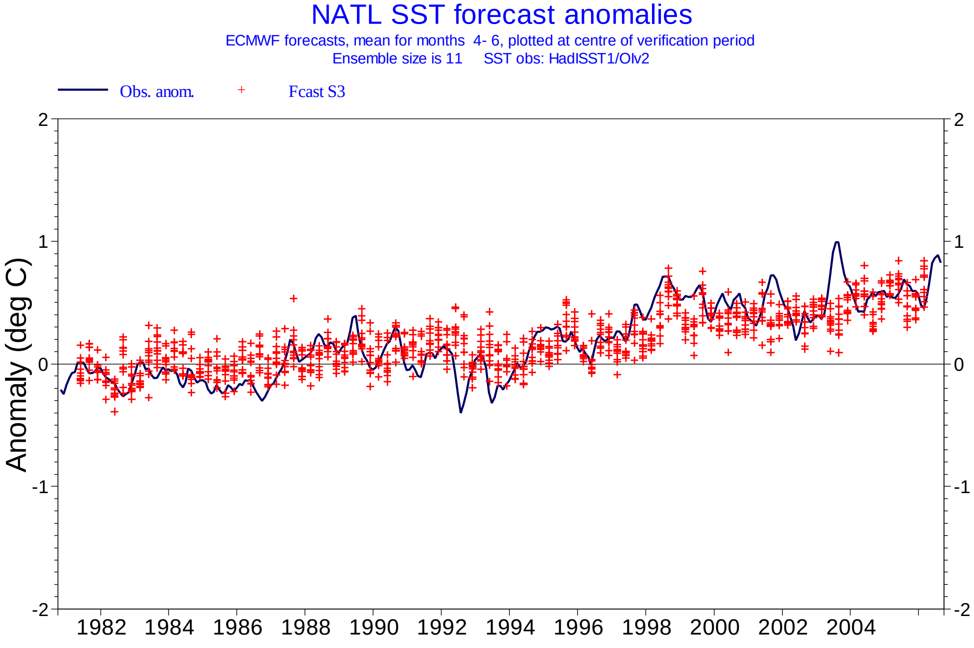


Four month lead

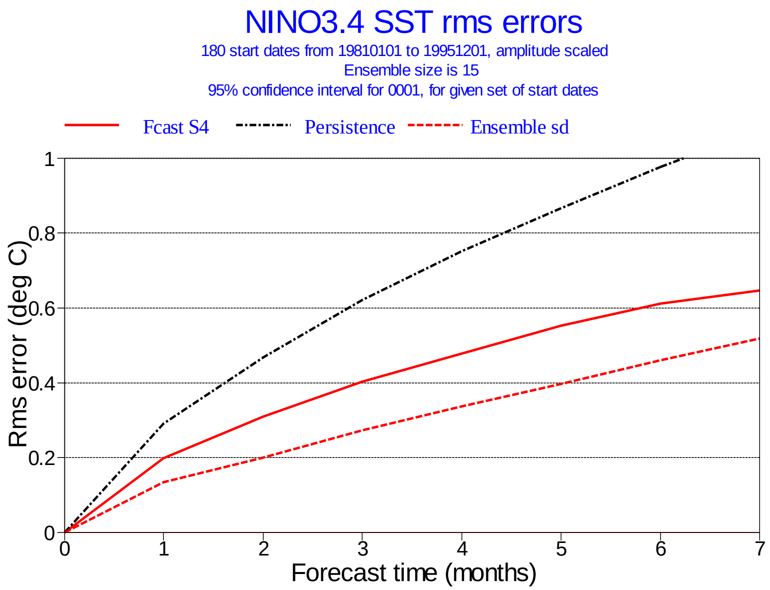
Capturing trends is important. Time-varying CO2 and other factors are important in this.

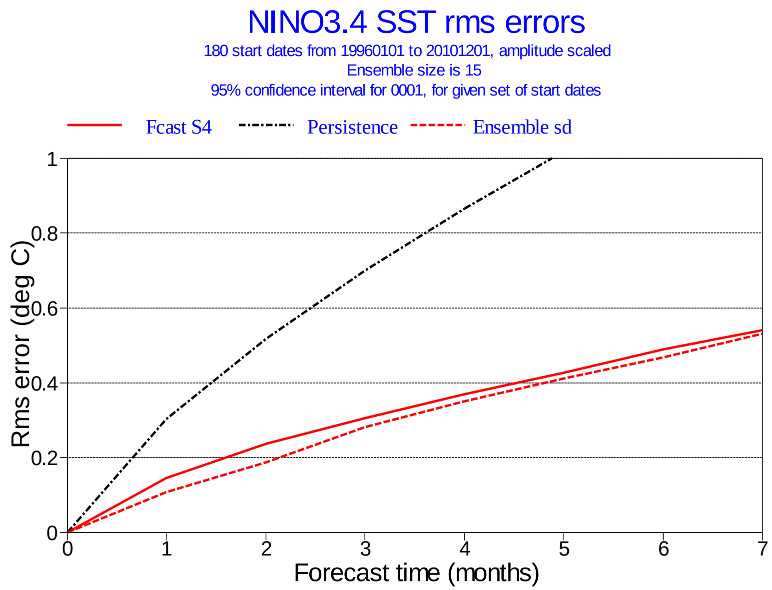
There is a strong link between seasonal prediction and decadal/ multi-decadal climate prediction.





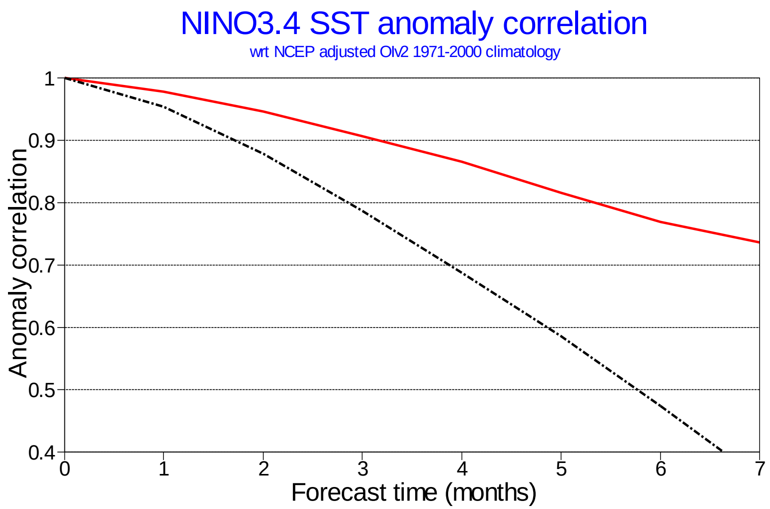
More recent ENSO forecasts are better ….

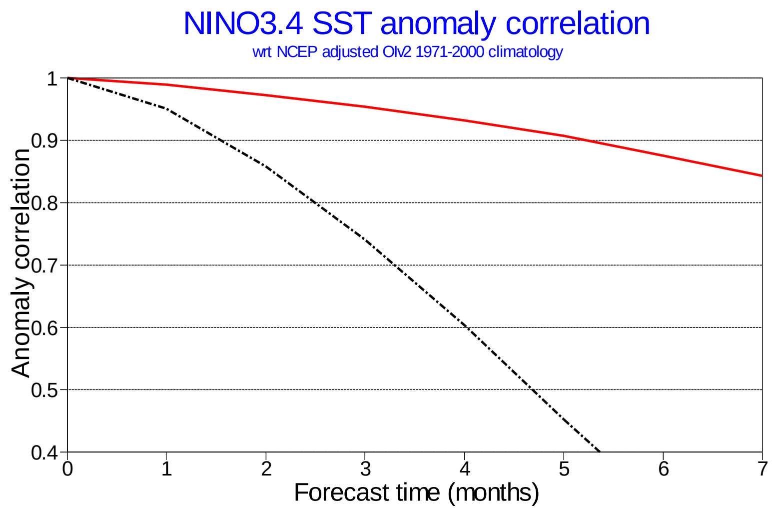




1981-1995

1996-2010





System 4 configuration

- **Real time forecasts:**
  - **51 member ensemble forecast to 7 months**
  - SST and atmos. perturbations added to each member

  - **15 member ensemble forecast to 13 months**
  - Designed to give an ‘outlook’ for ENSO
  - Only once per quarter (Feb, May, Aug and Nov starts)

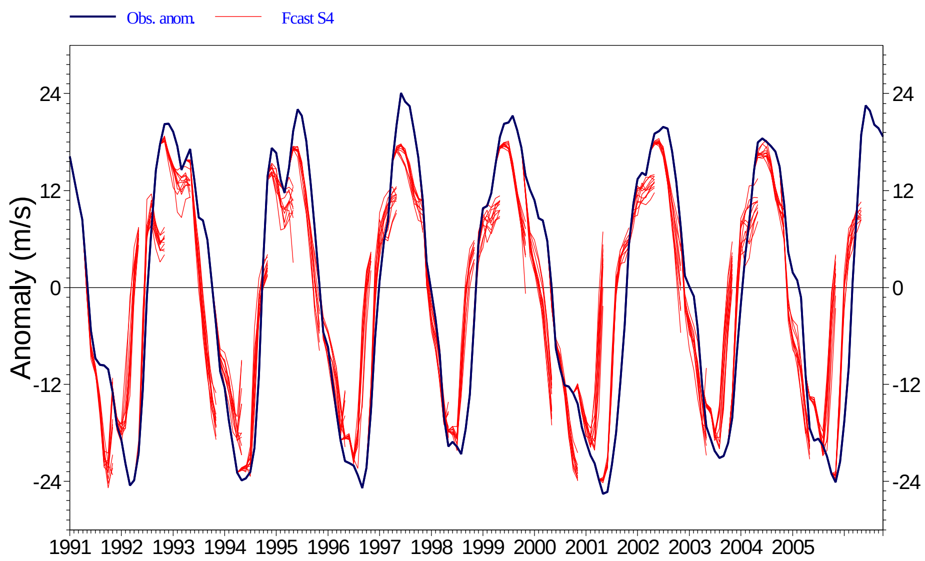
- **Back integrations from 1981-2010 (30 years)**
  - 15 member ensemble every month
  - 15 members extended to 13 months once per quarter

How many back integrations?

- **Back integrations dominate total cost of system**
  - System 4: 5400 back integrations (must be in first year)
  - 612 real-time integrations (per year)
- **Back integrations define model climate**
  - Need both climate mean and the pdf, latter needs large sample
  - May prefer to use a “recent” period (30 years? Or less??)
  - System 2 had a 75 member “climate”, S3 had 275, S4 has 450.
  - Sampling is basically OK
- **Back integrations provide information on skill**
  - A forecast cannot be used unless we know (or assume) its level of skill
  - Observations have only 1 member, so large ensembles are less helpful than large numbers of cases.
  - Care needed e.g. to estimate skill of 51 member ensemble based on past performance of 15 member ensemble
  - For regions of high signal/noise, System 4 gives adequate skill estimates
  - For regions of low signal/noise (eg <= 0.5), need hundreds of years

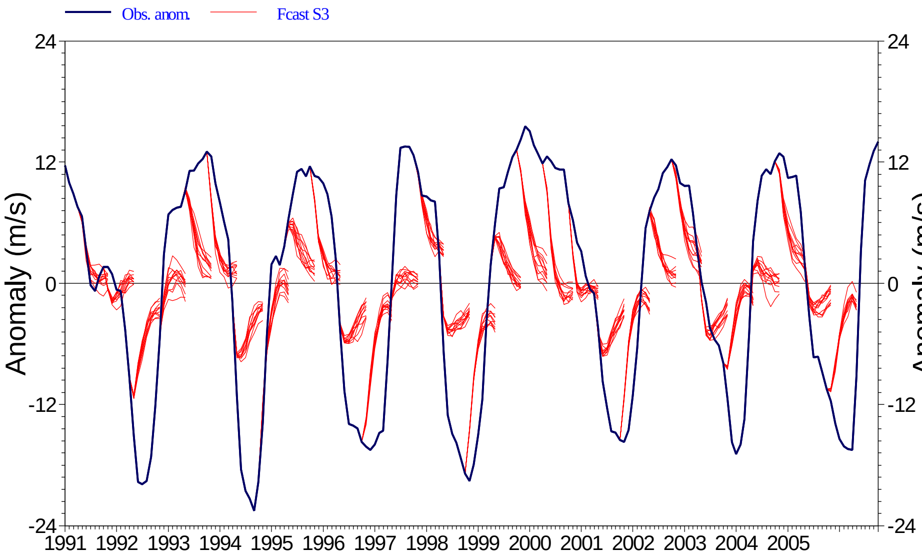
QBO

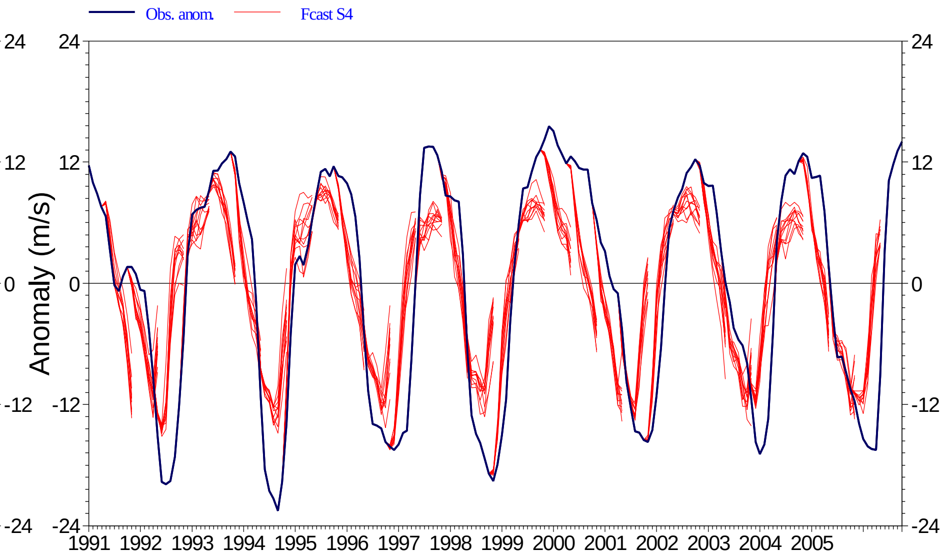
System 4



30hPa

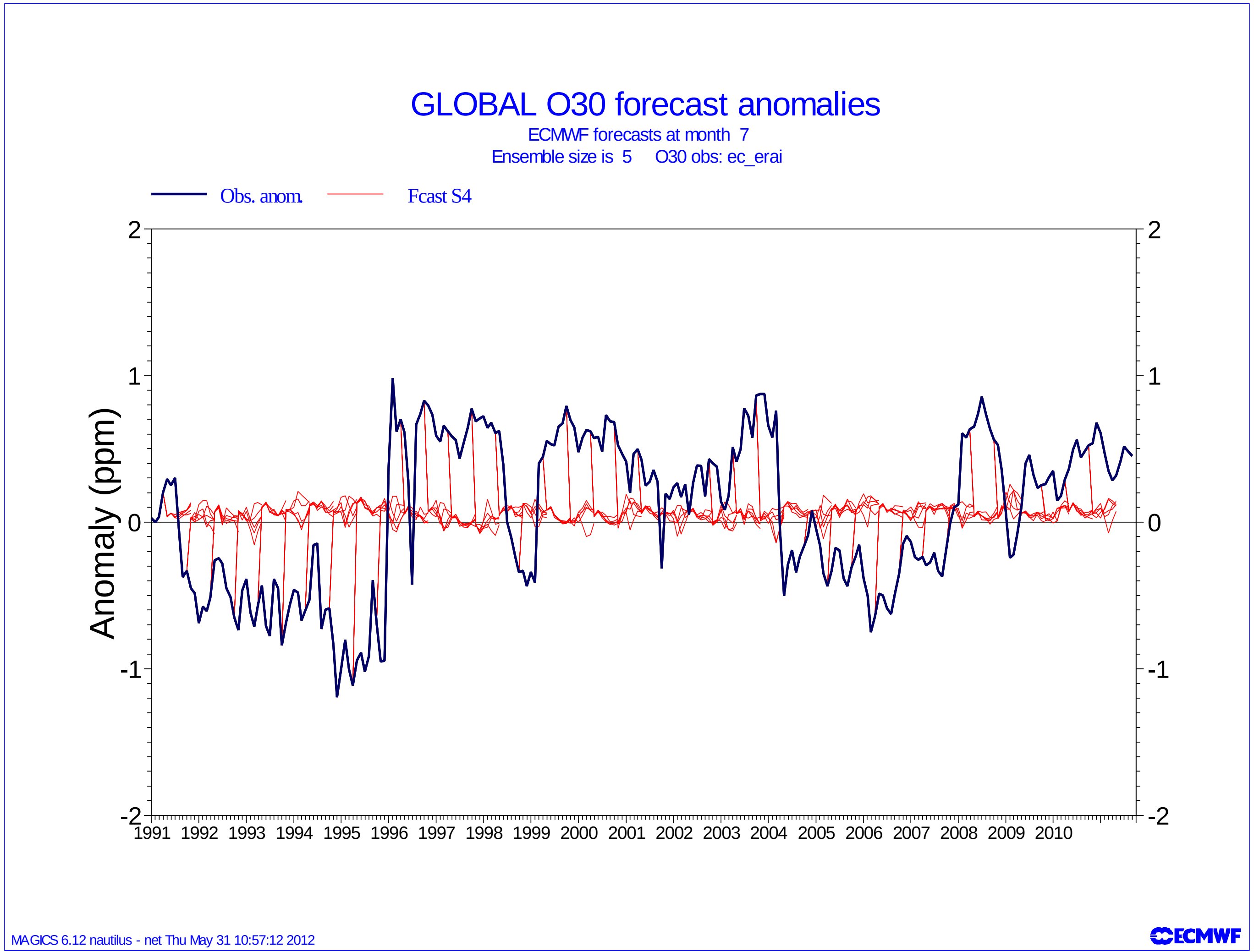
System 3



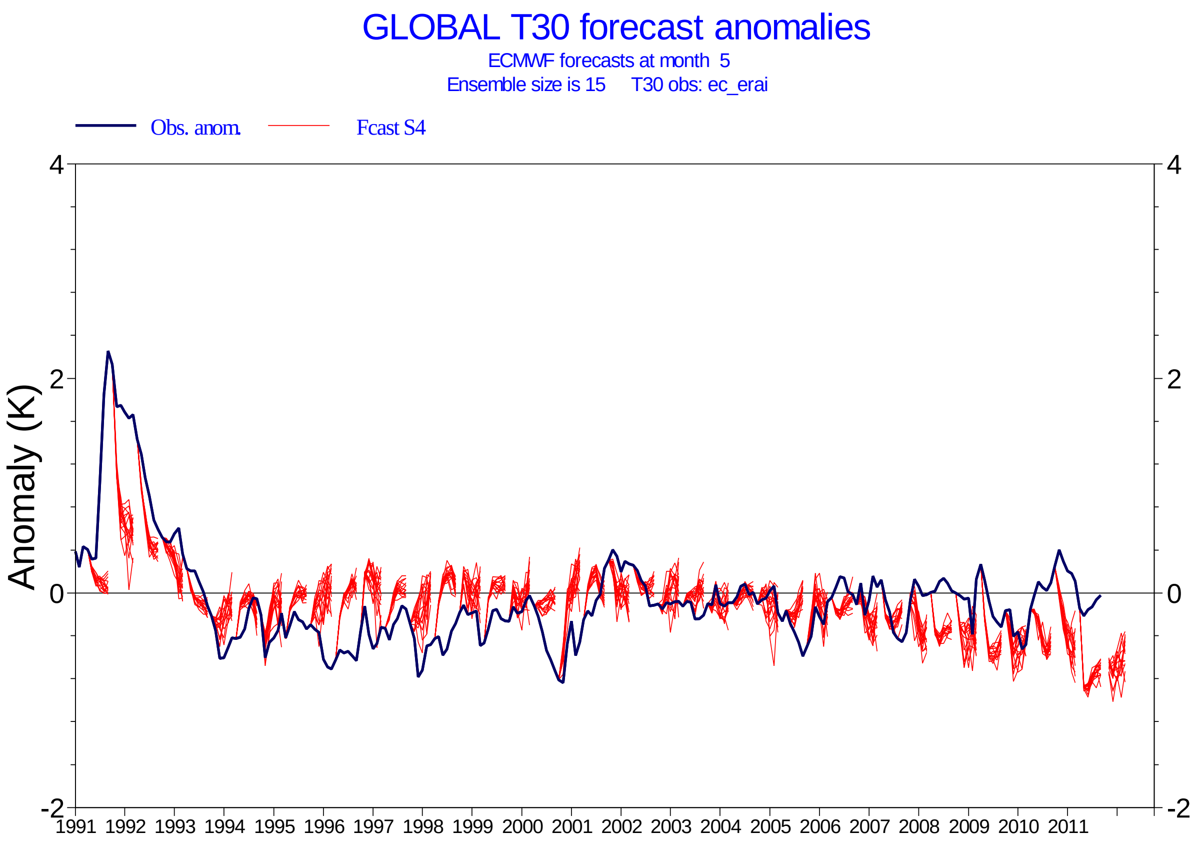


50hPa

Problematic ozone analyses

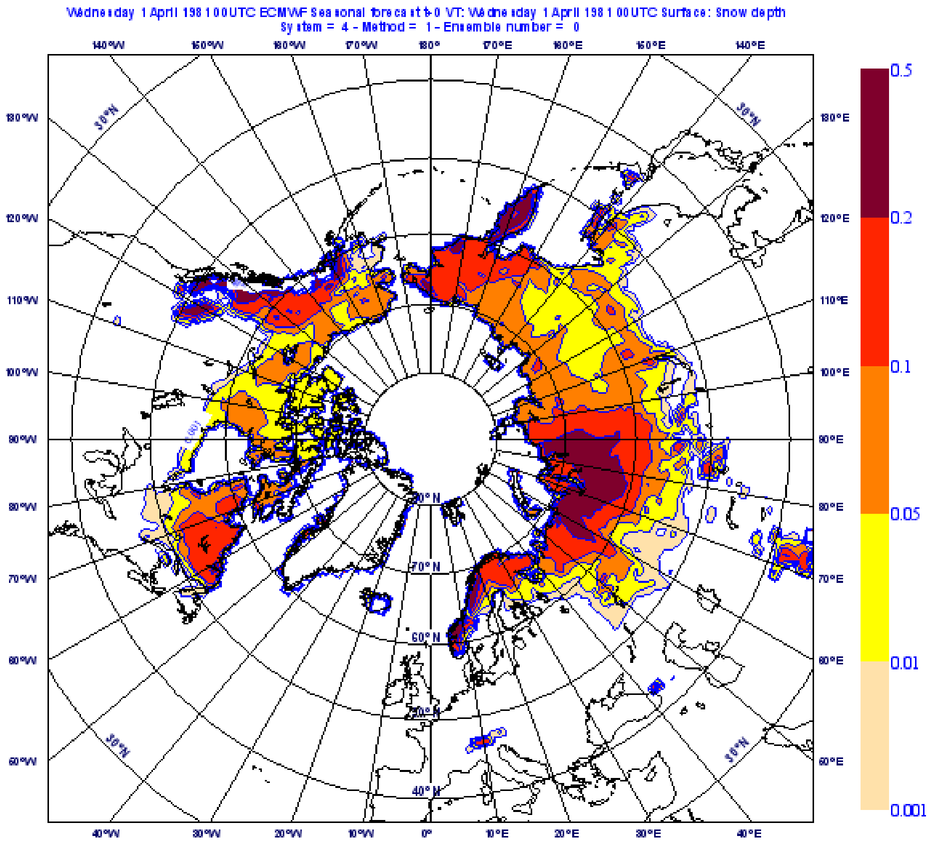


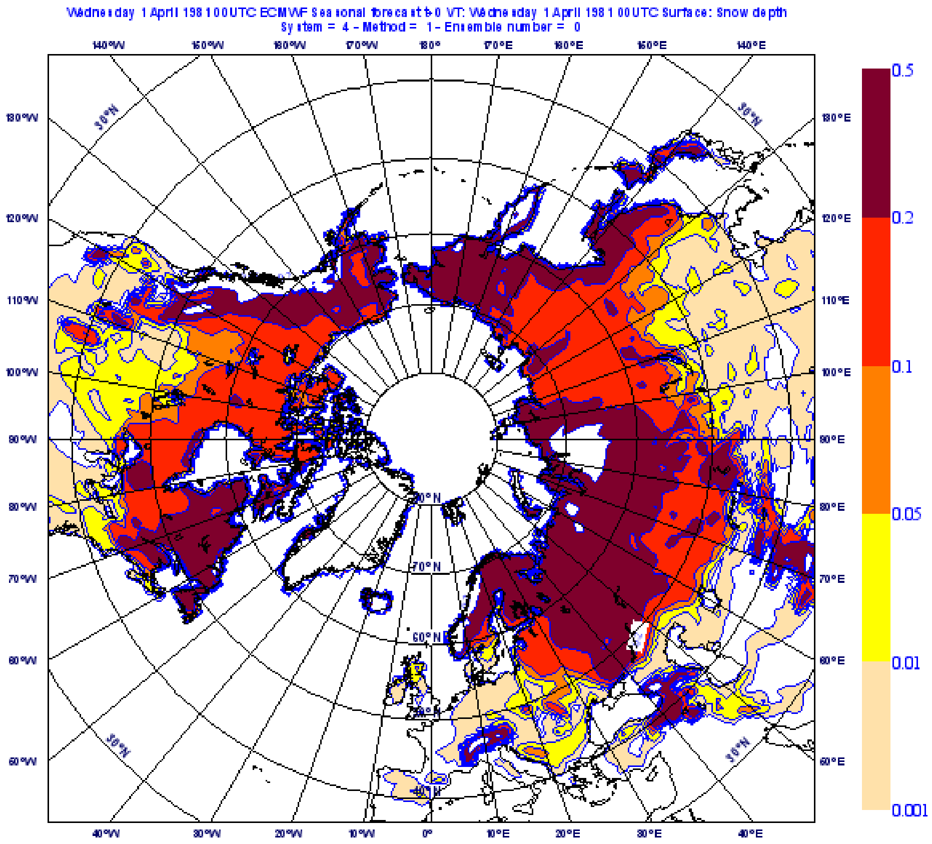
Stratospheric trends



Stratospheric temperature trend problem. This is due to an erroneous trend in initial conditions of stratospheric water vapour.

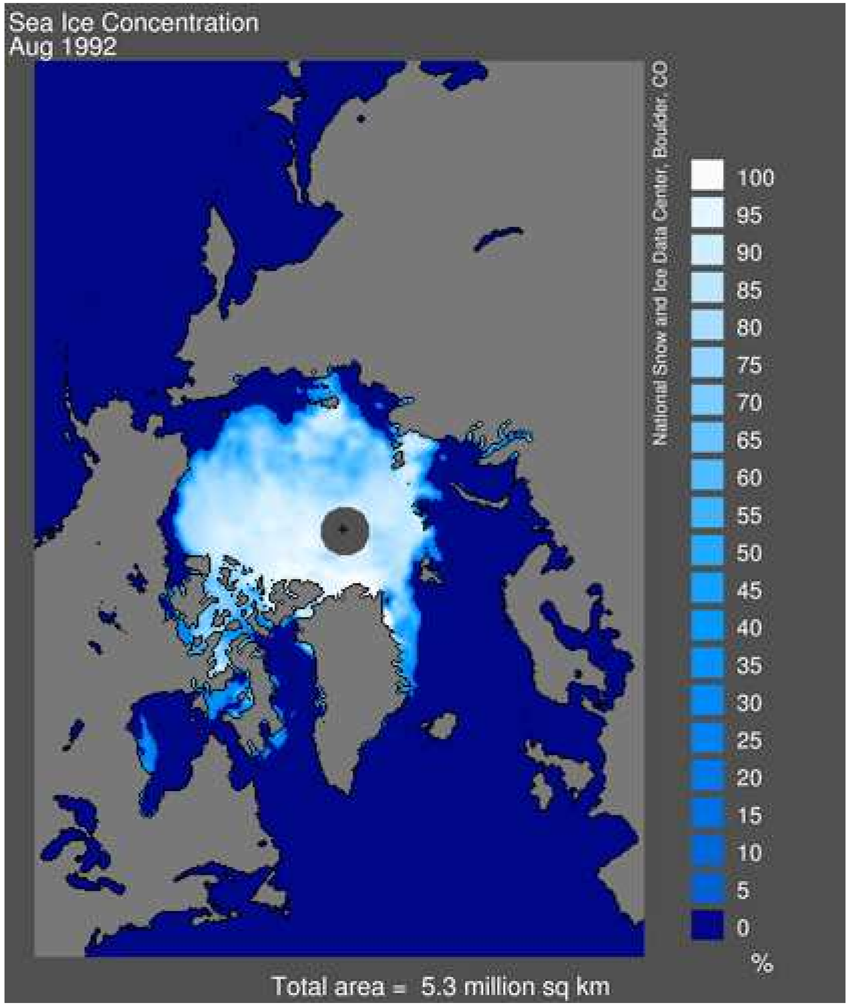
Land surface

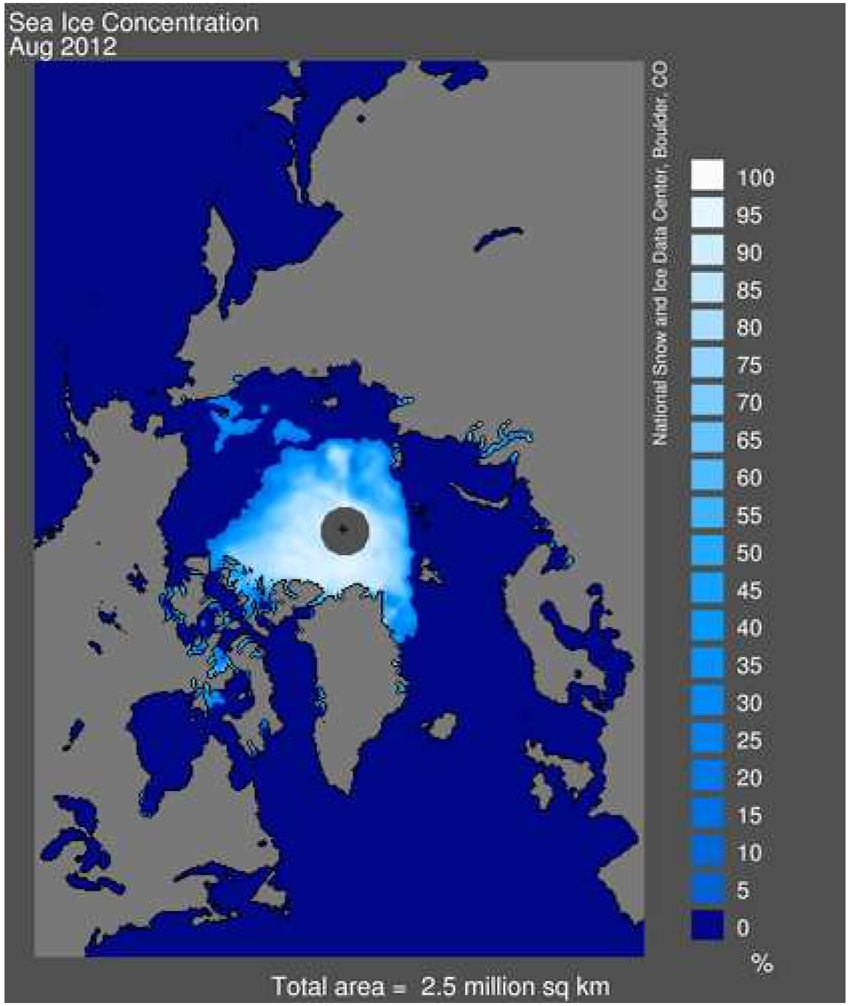




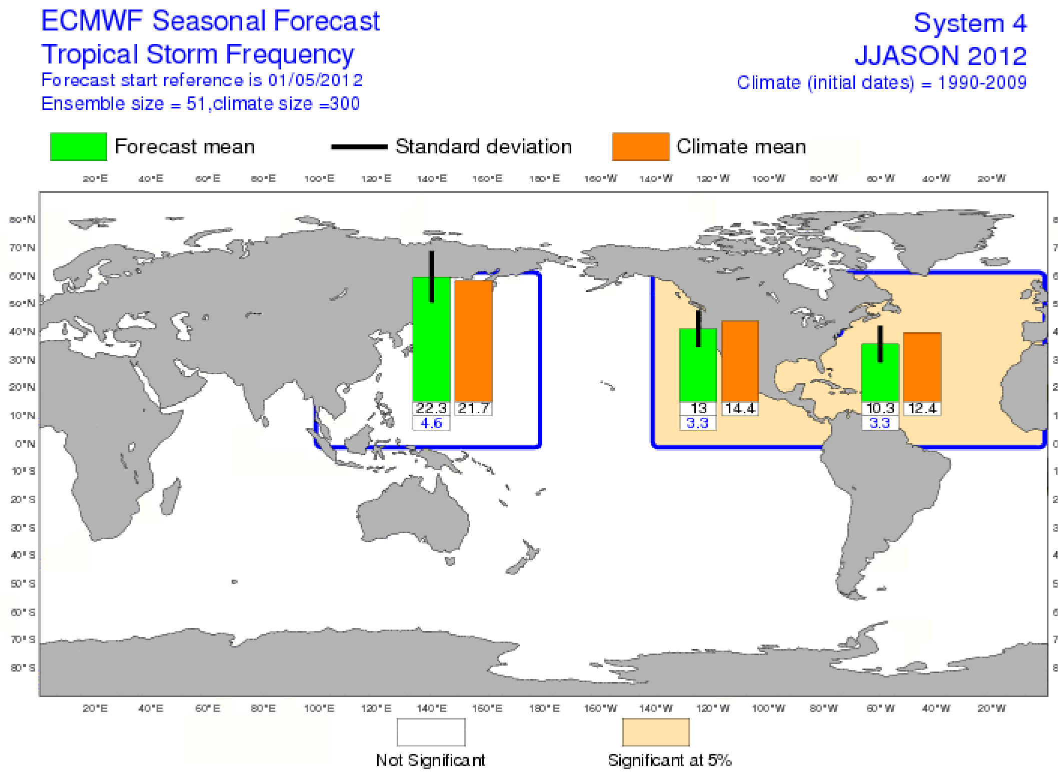
Snow depth limits, 1st April

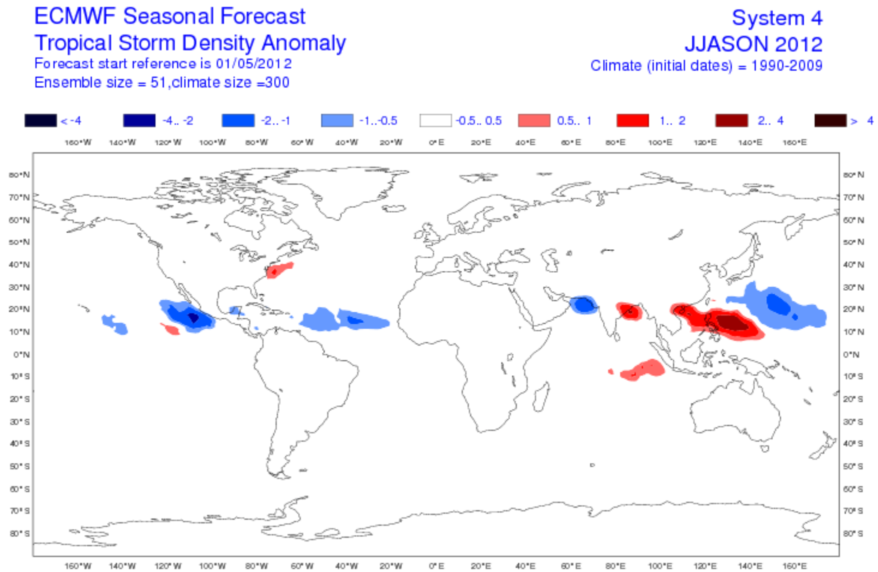
Sea ice

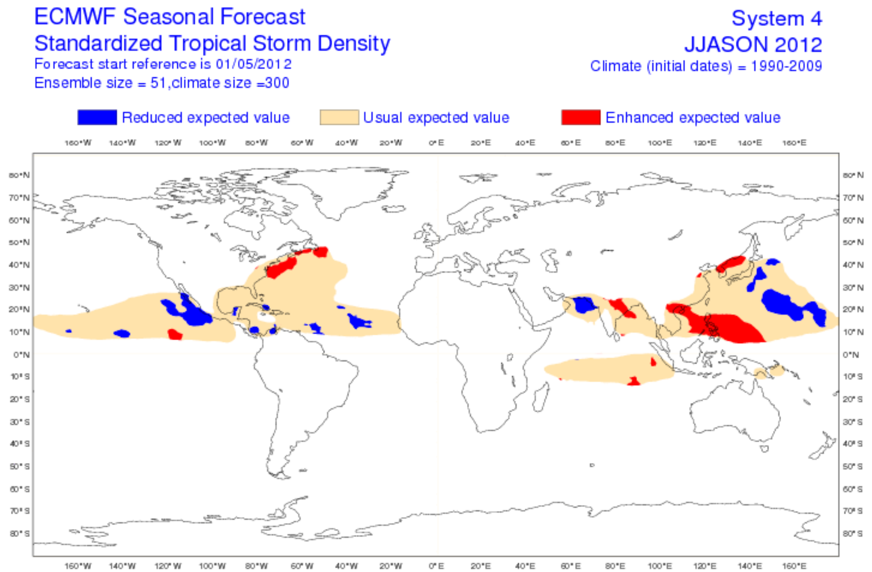

Sea Ice Concentration
Aug 1992
National Snow and Ice Data Center, Boulder, CO
100
95
90
85
80
75
70
65
60
55
50
45
40
35
30
25
20
15
10
5
0
%
Total area = 5.3 million sq km


Sea Ice Concentration
Aug 2012
National Snow and Ice Data Center, Boulder, CO
100
95
90
85
80
75
70
65
60
55
50
45
40
35
30
25
20
15
10
5
0
%
Total area = 2.5 million sq km

Tropical storm forecasts







Model errors are still serious ...

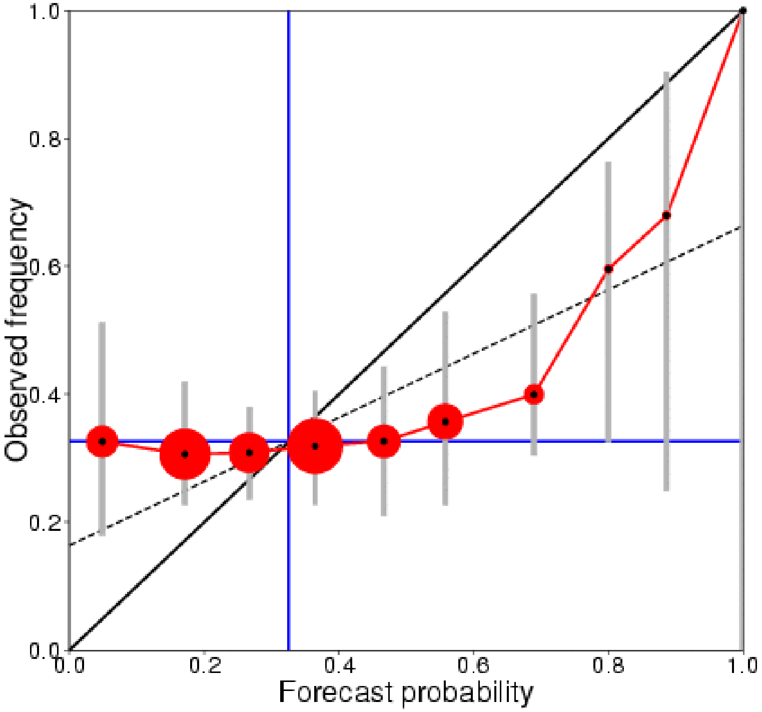
- **Models have errors other than mean bias**
  - Eg weak wind and SST variability in System 2
  - Past models underestimated MJO activity (S4 better)
  - Suspected too-weak teleconnections to mid-latitudes
- **Mean state errors interact with model variability**
  - Nino 4 region is very sensitive (cold tongue/warm pool boundary)
  - Atlantic variability suppressed if mean state is badly wrong
- **Forecast errors are often larger than they should be**
  - With respect to internal variability estimates and (occasionally) other prediction systems
  - Reliability of probabilistic forecasts is often not particularly high (S4 better)

Recent Research

S4 extended hindcast set

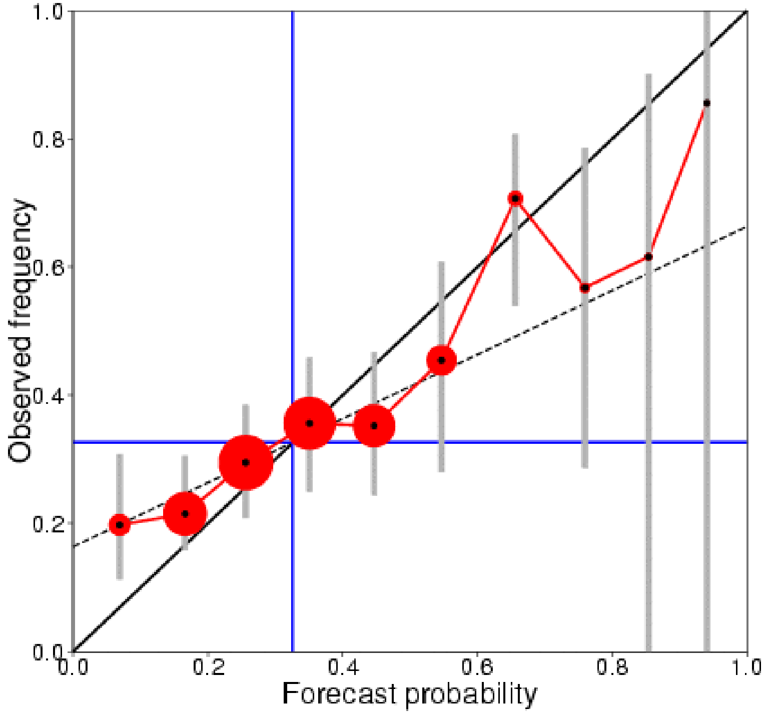
15 members

DJF Europe T2m>upper tercile
Re-forecasts from 1 Nov, 1981-2010
Reliability score: 0.902
ROC skill score: 0.06



51 members

DJF Europe T2m>upper tercile
Re-forecasts from 1 Nov, 1981-2010
Reliability score: 0.981
ROC skill score: 0.22

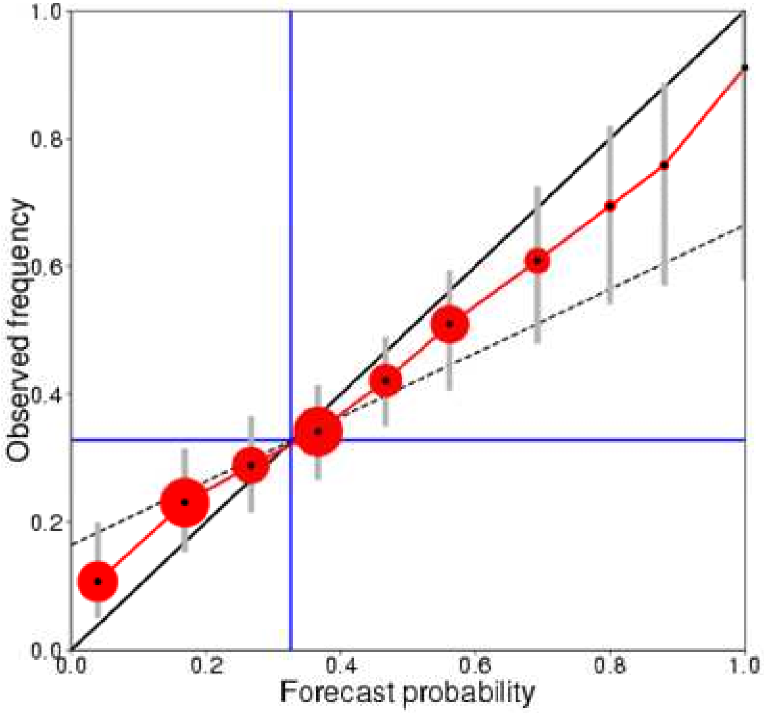


(Figures from Susanna Corti)

S4 extended hindcast set

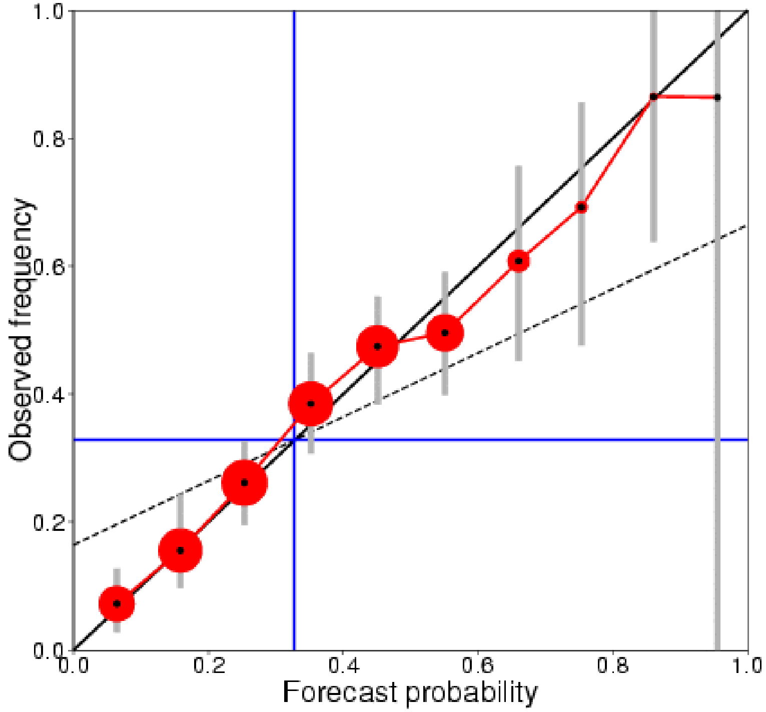
15 members

JJA Europe T2m>upper tercile
Re-forecasts from 1 May, 1981-2010
Reliability score: 0.987
ROC skill score: 0.38



51 members

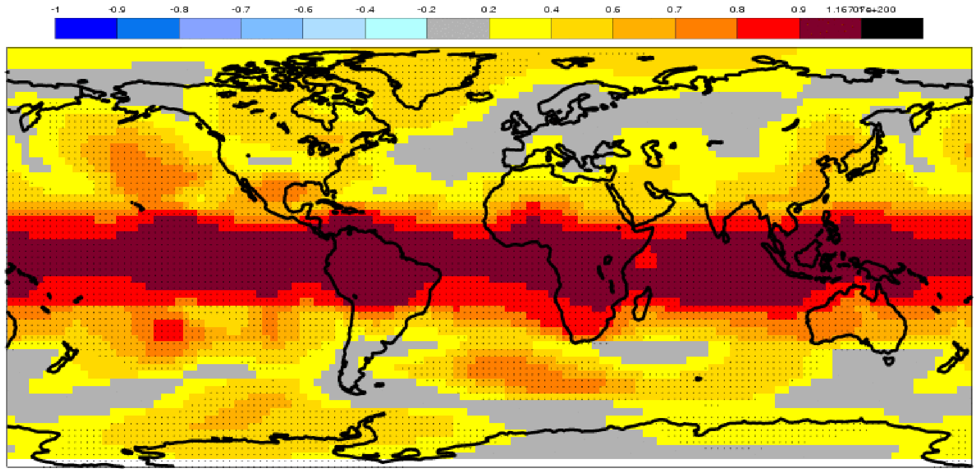
JJA Europe T2m>upper tercile
Re-forecasts from 1 May, 1981-2010
Reliability score: 0.996
ROC skill score: 0.43



(Figures from Susanna Corti)

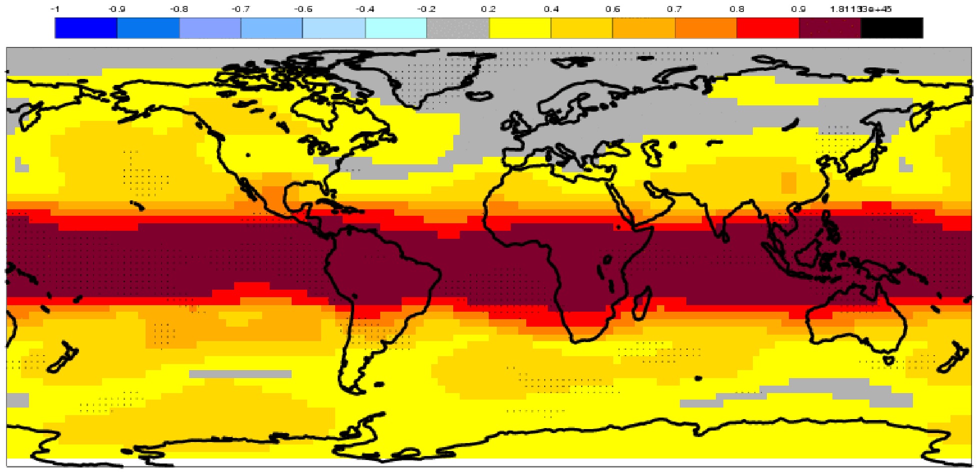
S4 ACC
DJF Z500

Anomaly Correlation Coefficient for ECMWF S4 with 51 ensemble members
500 hPa geopotential height
Hindcast period 1981-2010 with start in November average over months 2 to 4
Black dots for values significantly different from zero with 95% confidence (1000 samples)



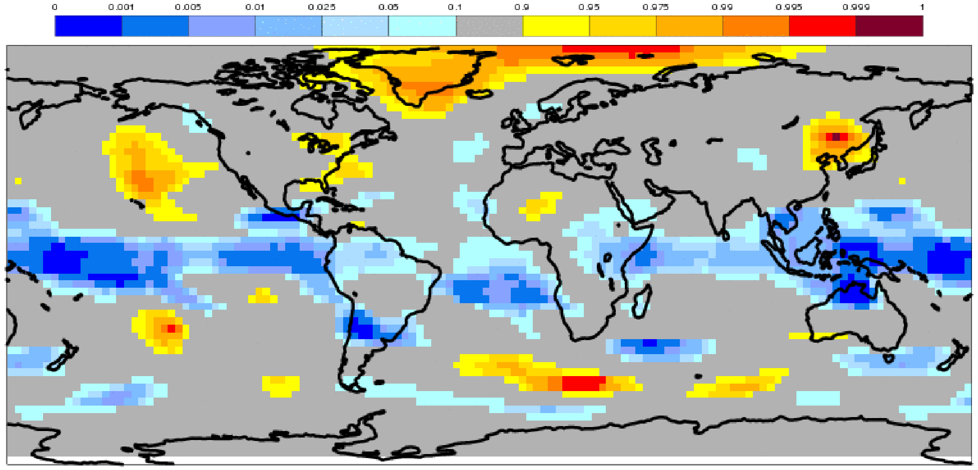
S4 ACC perfect model limit

Perfect-model Anomaly Correlation Coefficient for ECMWF S4 with 51 ensemble members
500 hPa geopotential height
Hindcast period 1981-2010 with start in November average over months 2 to 4
Black dots where perfect model assumption is violated with 95% confidence (1000 samples)

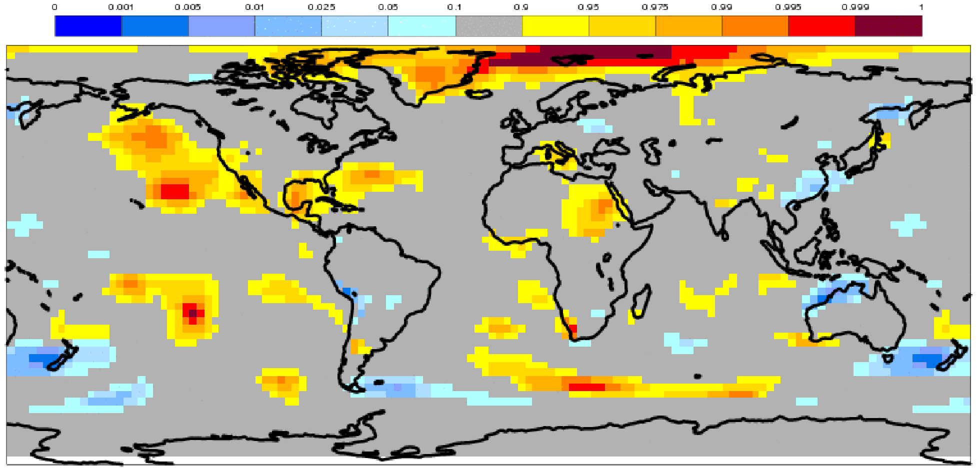


Local p-value for perfect model

p-value for observed ACC, assuming perfect model for ECMWF S4 with 51 ensemble members
500 hPa geopotential height
Hindcast period 1981-2010 with start in November average over months 2 to 4



p-value for observed ACC, assuming perfect model for ECMWF S4 with 51 ensemble members
Mean sea level pressure
Hindcast period 1981-2010 with start in November average over months 2 to 4



Indistinguishable from perfect
Worse than perfect
Better than perfect

ACC=0.61

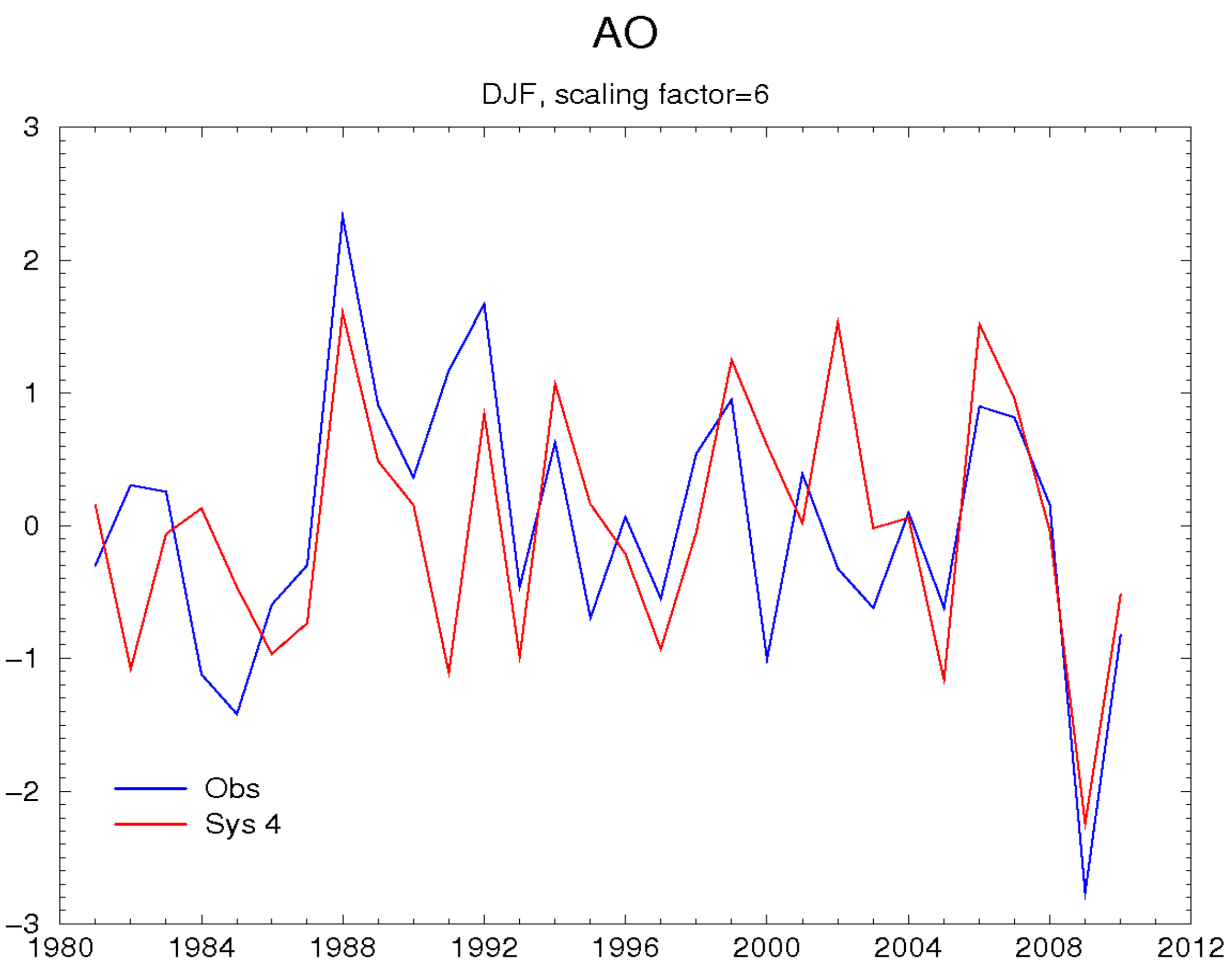
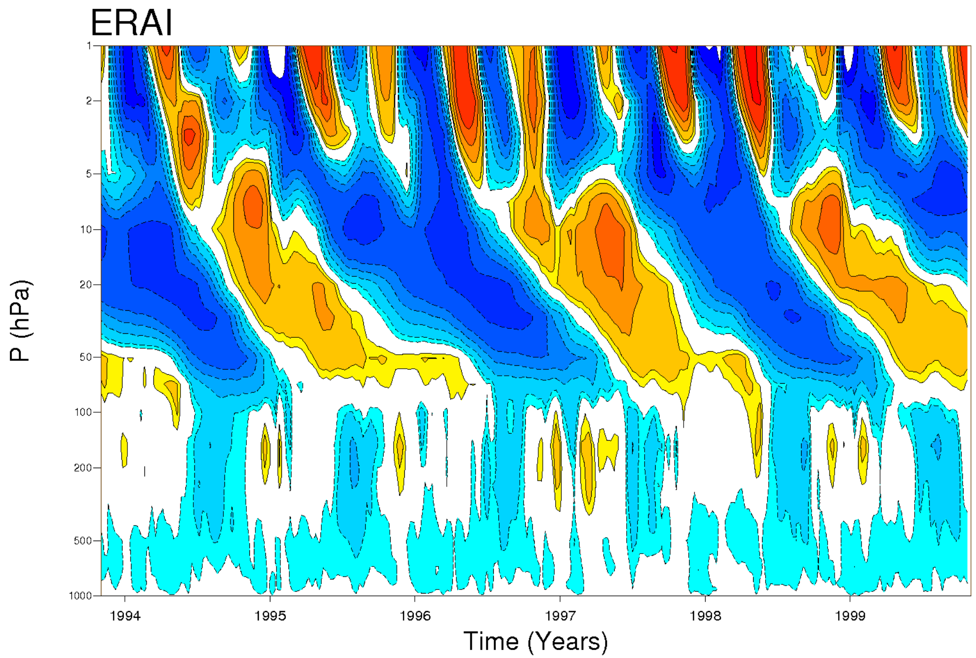
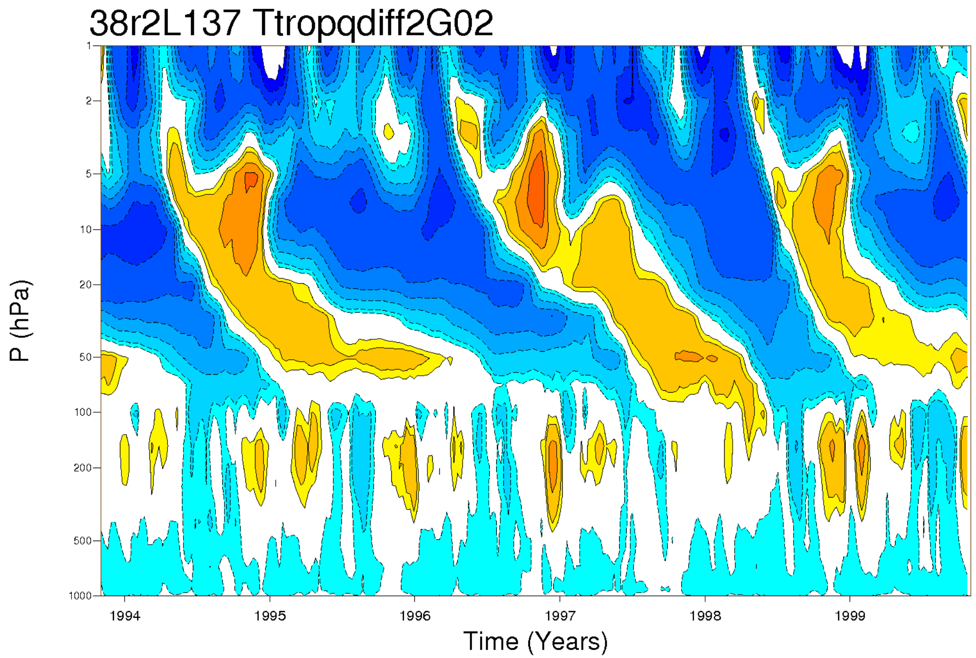


Fig. 1. The Arctic Oscillation Index for DJF, as analysed from ERAI (blue) and as predicted by the S4 ensemble mean from the 1st November (red). The S4 ensemble mean is **scaled by a factor of 6** to be of comparable amplitude to the observed index.

QBO

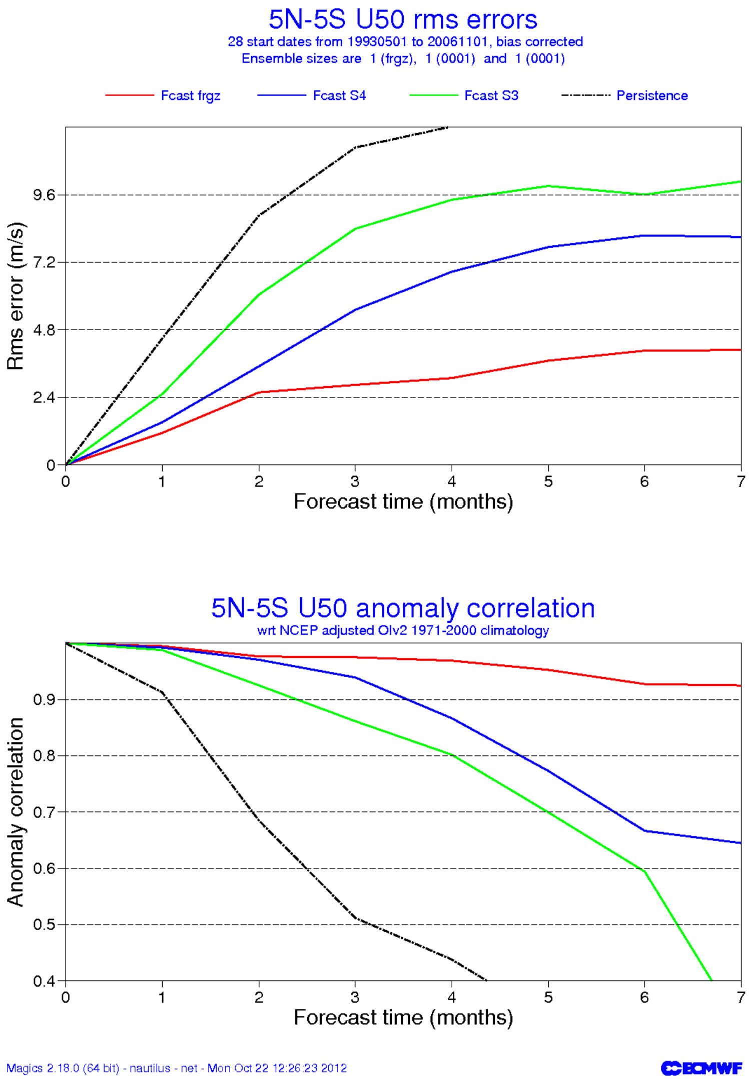
A big reduction in vertical diffusion, and a further tuning of non-orographic GWD, has given a big additional improvement in the QBO compared to S4.





Period and downward penetration match observations
Semi-annual oscillation still poorly represented

QBO forecasts



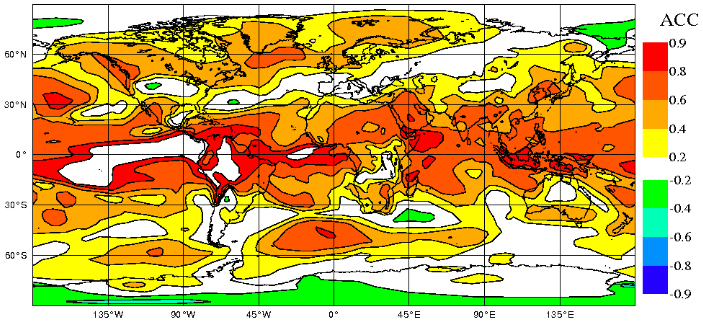
S3

S4

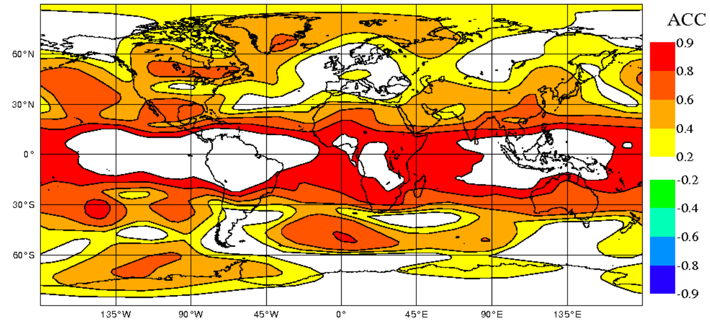
New

NH winter forecasts: vertical diffusion

T850 Anom. correlation fuhg(101)-ERA-Int 1981-2010DJF
Global z-mean acc: 0.513 NH:0.351 TR:0.662 SH:0.28

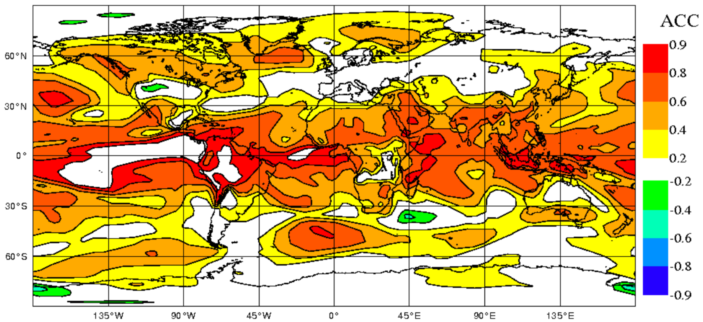


Z500 Anom. correlation fuhg(101)-ERA-Int 1981-2010DJF
Global z-mean acc: 0.666 NH:0.371 TR:0.829 SH:0.403

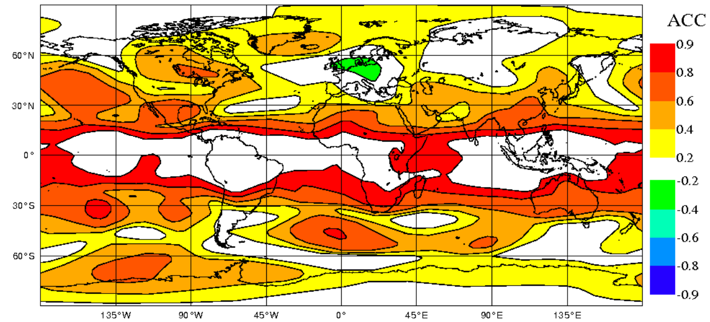


0.371

T850 Anom. correlation fulf(101)-ERA-Int 1981-2010DJF
Global z-mean acc: 0.503 NH:0.309 TR:0.66 SH:0.283

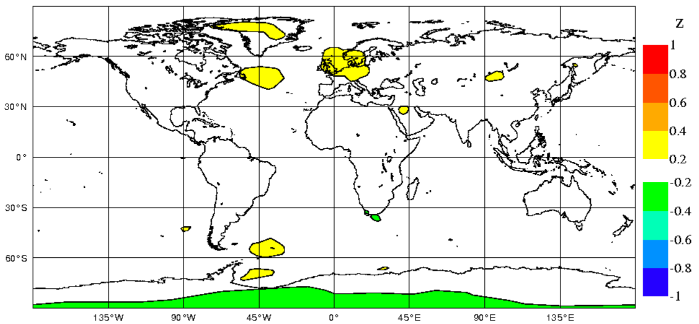


Z500 Anom. correlation fulf(101)-ERA-Int 1981-2010DJF
Global z-mean acc: 0.664 NH:0.319 TR:0.832 SH:0.424

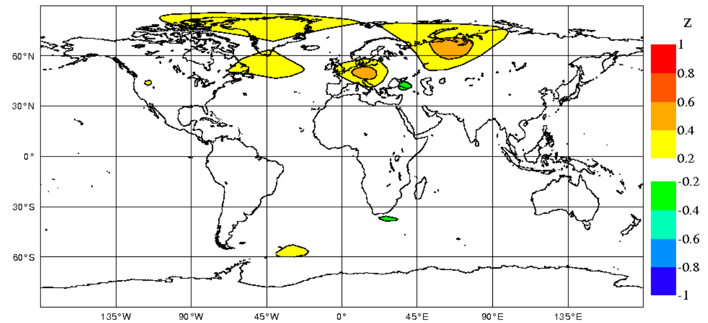


0.319

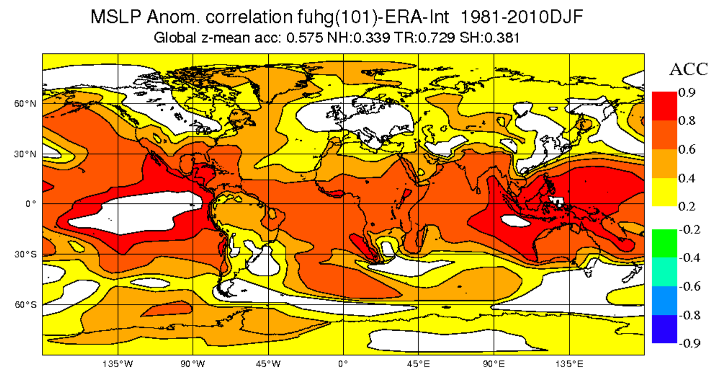
Fisher z transform diff fuhg(101)-fulf(101) 1981-2010DJF
sigma: 0.272 mean: 0.0133

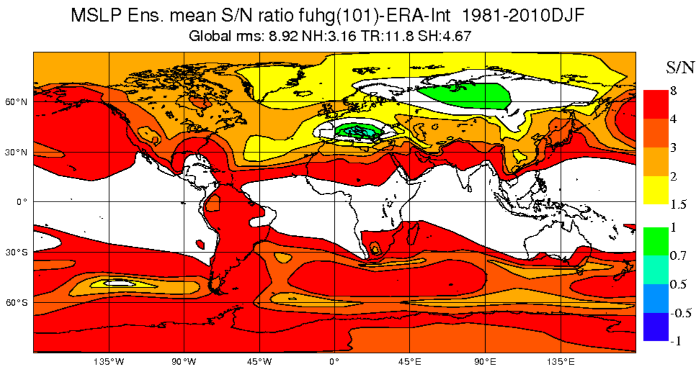


Fisher z transform diff fuhg(101)-fulf(101) 1981-2010DJF
sigma: 0.272 mean: 0.004

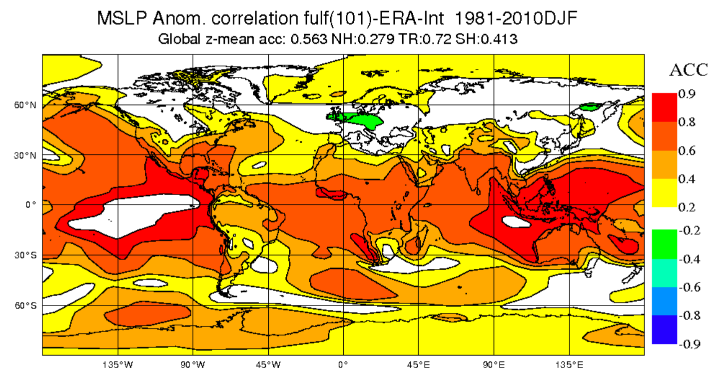


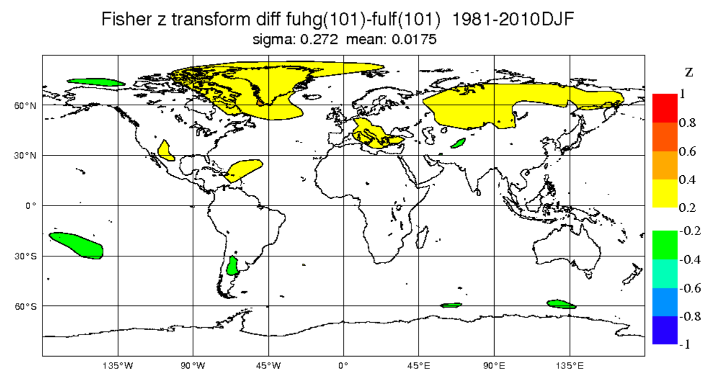
NH winter forecasts





Even with 101 members, ensemble mean signal not always well defined





Conclusions

- **Models are improving**
  - Gradual but continuous improvement in scores
  - Reliability can be high in many situations
- **Forecast systems still have deficiencies**
  - Need calibration, and often cannot be trusted at face value
  - Some issues may affect real-time forecasts more than re-forecasts
- **Further improvements lie ahead**
  - Research results suggesting that previous estimates of predictability limits might be *wrong.*
  - Hard work needed to improve models and capture new sources of predictability.