An extended re-forecast set for ECMWF system 4

in the context of EUROSIP

Tim Stockdale

Acknowledgements:

Magdalena Balmaseda, Susanna Corti, Laura Ferranti, Kristian Mogensen, Franco Molteni, Frederic Vitart, Antje Weisheimer

Outline

- **Operational seasonal prediction**

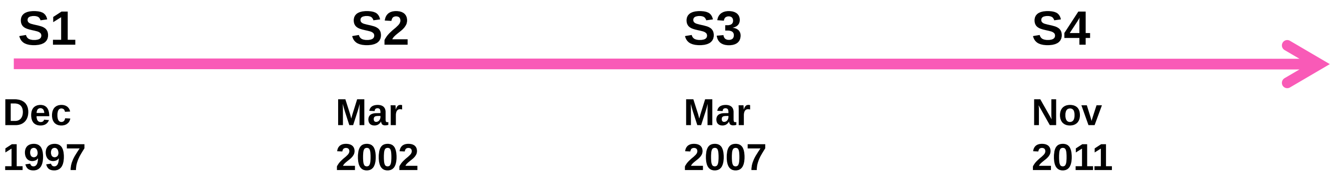
- **ECMWF S4**

- **An extended re-forecast set for S4**
  - Statistical testing
  - Why “better than perfect” is not what we want ….

- **EUROSIP – a multi-model collaboration**

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



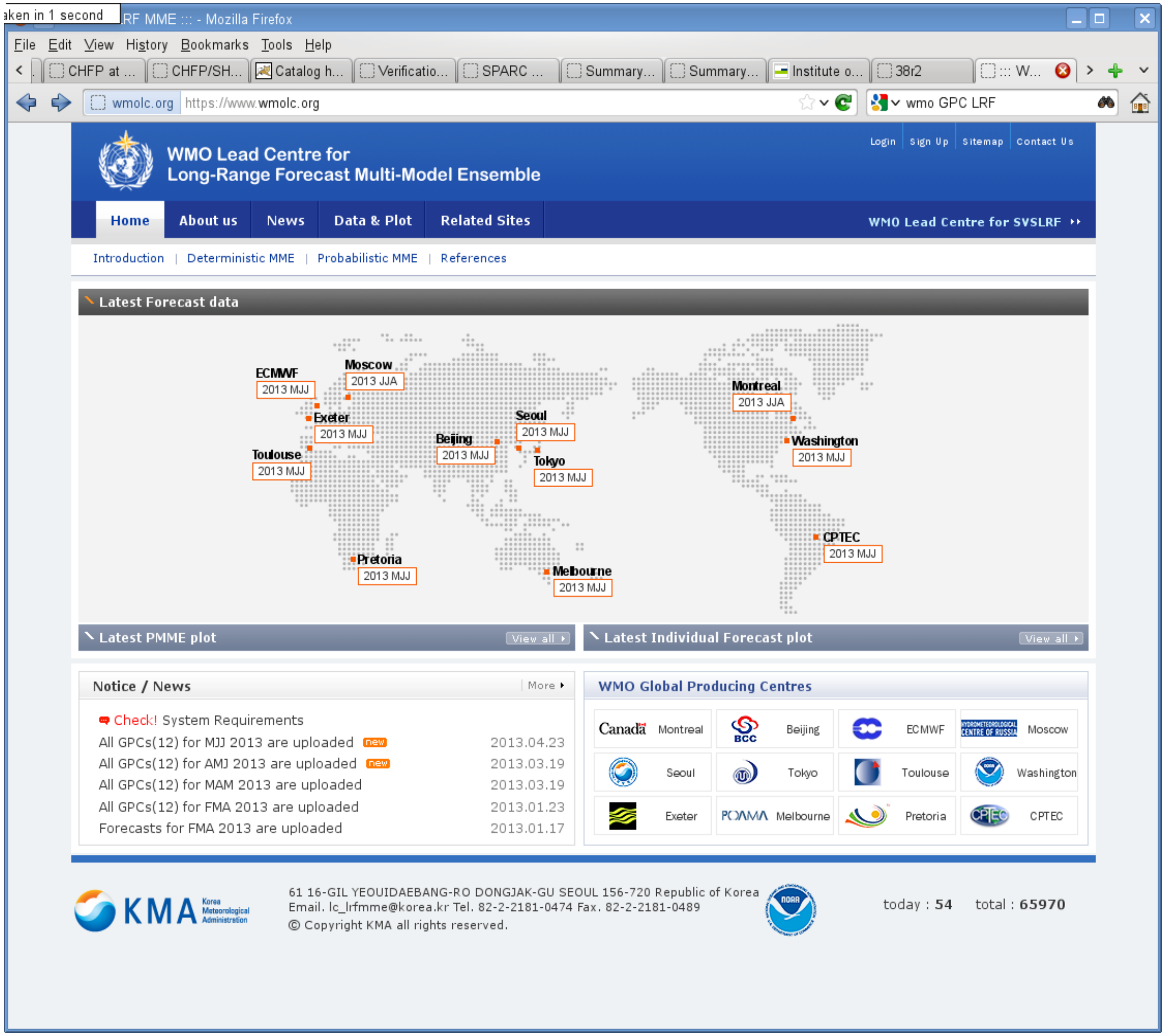
S1 Dec 1997

S2 Mar 2002

S3 Mar 2007

S4 Nov 2011

WMO-designated “Global Producing Centres”


aken in 1 second
RF MME ::: - Mozilla Firefox
File Edit View History Bookmarks Tools Help
CHFP at ... | CHFP/SH... | Catalog h... | Verificatio... | SPARC ... | Summary... | Summary... | Institute o... | 38r2 | ::: W...
wmolc.org https://www.wmolc.org
wmo GPC LRF
Login | Sign Up | Sitemap | Contact Us
Home | About us | News | Data & Plot | Related Sites
WMO Lead Centre for SVSLRF
Introduction | Deterministic MME | Probabilistic MME | References
Latest Forecast data
ECMWF 2013 MJJ
Moscow 2013 JJA
Exeter 2013 MJJ
Toulouse 2013 MJJ
Beijing 2013 MJJ
Seoul 2013 MJJ
Tokyo 2013 MJJ
Pretoria 2013 MJJ
Melbourne 2013 MJJ
Montreal 2013 JJA
Washington 2013 MJJ
CPTEC 2013 MJJ
Latest PMME plot View all
Latest Individual Forecast plot View all
More

WMO Lead Centre for
Long-Range Forecast Multi-Model Ensemble

Notice / News

Check! System Requirements

All GPCs(12) for MJJ 2013 are uploaded new	2013.04.23
All GPCs(12) for AMJ 2013 are uploaded new	2013.03.19
All GPCs(12) for MAM 2013 are uploaded	2013.03.19
All GPCs(12) for FMA 2013 are uploaded	2013.01.23
Forecasts for FMA 2013 are uploaded	2013.01.17

WMO Global Producing Centres

Canada Montreal	BCC Beijing	ECMWF	HYDROMETEOROLOGICAL CENTRE OF RUSSIA Moscow
Seoul	Tokyo	Toulouse	Washington
Exeter	POAMA Melbourne	Pretoria	CPTEC CPTEC

KMA Korea Meteorological Administration

61 16-GIL YEOUIDAEBANG-RO DONGJAK-GU SEOUL 156-720 Republic of Korea
Email. lc_lrfmme@korea.kr Tel. 82-2-2181-0474 Fax. 82-2-2181-0489
© Copyright KMA all rights reserved.

today : **54** total : **65970**

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

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

  - **51 members** for Feb/May/Aug/Nov starts

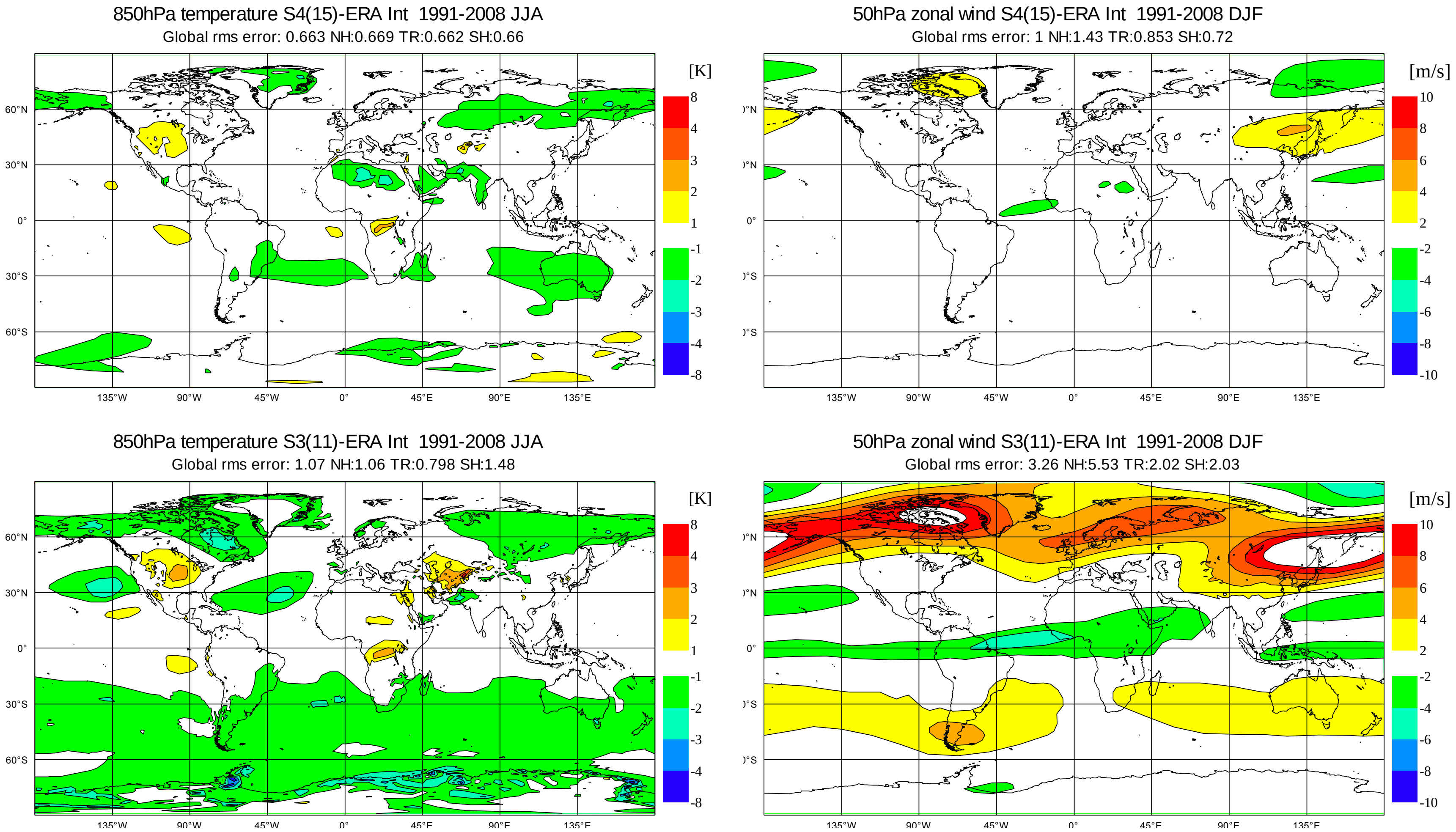
>> Data now available on CHFP server <<

Reduced mean state errors

T850 U50

S4

S3



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

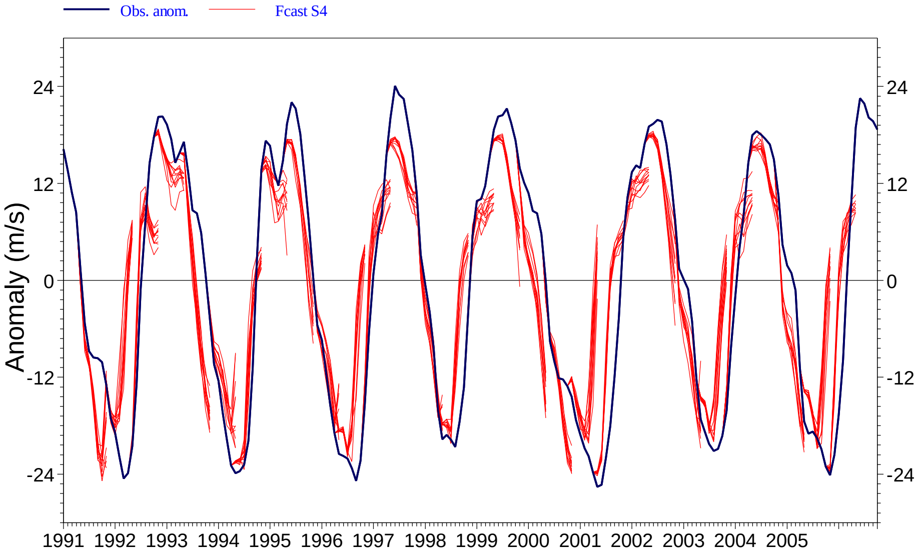
50hPa zonal wind S4(15)-ERA Int 1991-2008 DJF
Global rms error: 1 NH:1.43 TR:0.853 SH:0.72

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

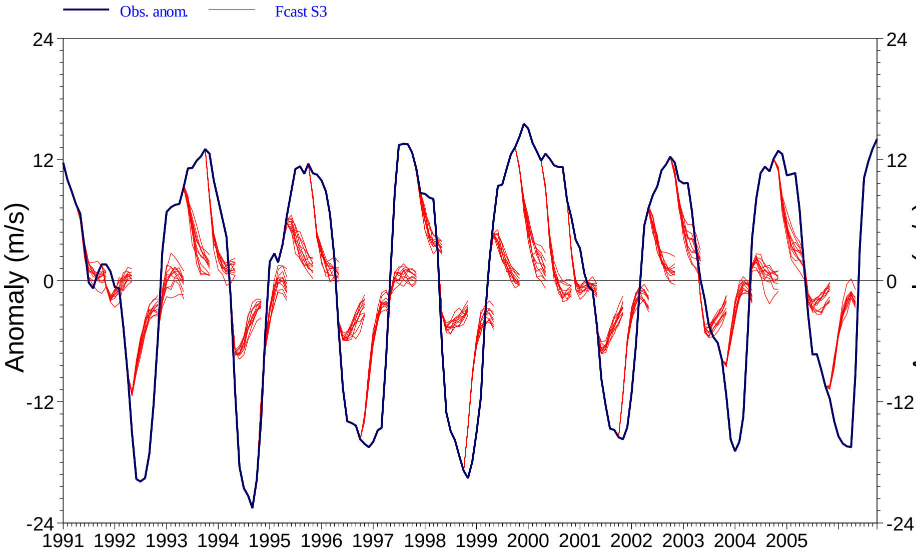
QBO

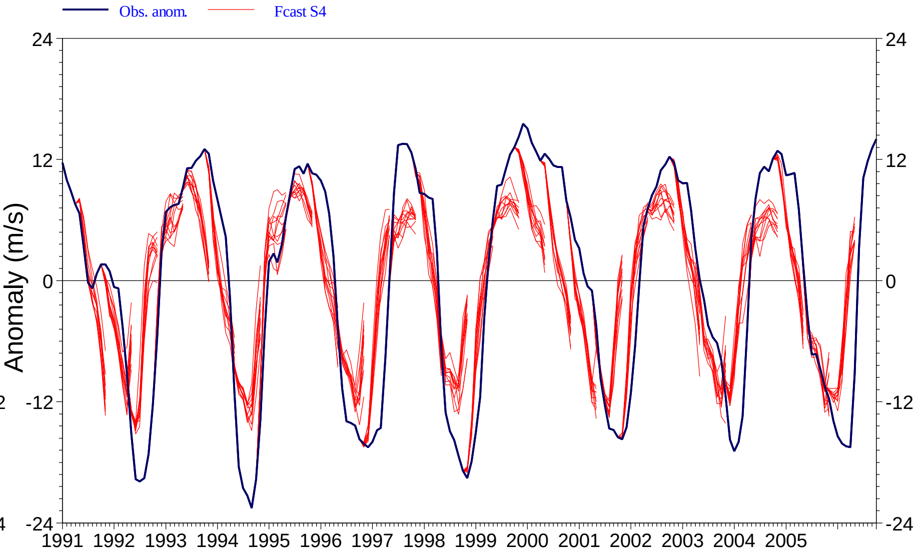
System 4



30hPa

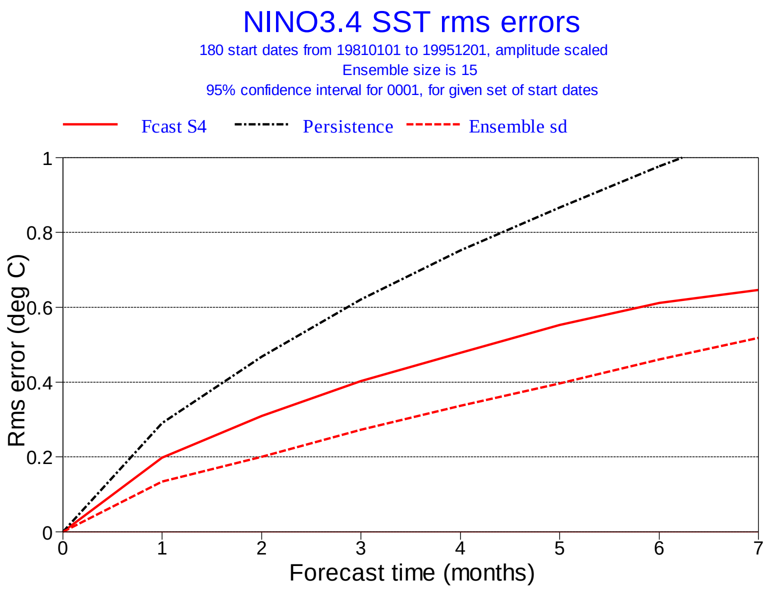
System 3

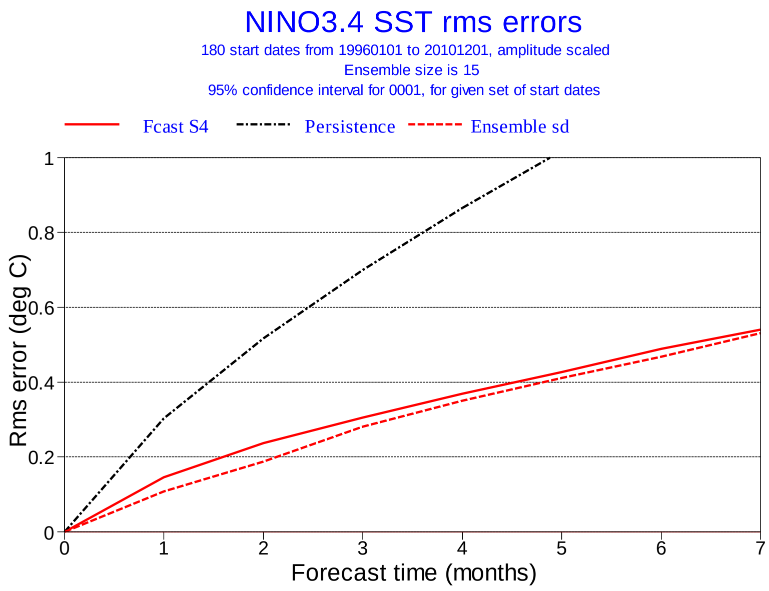




50hPa

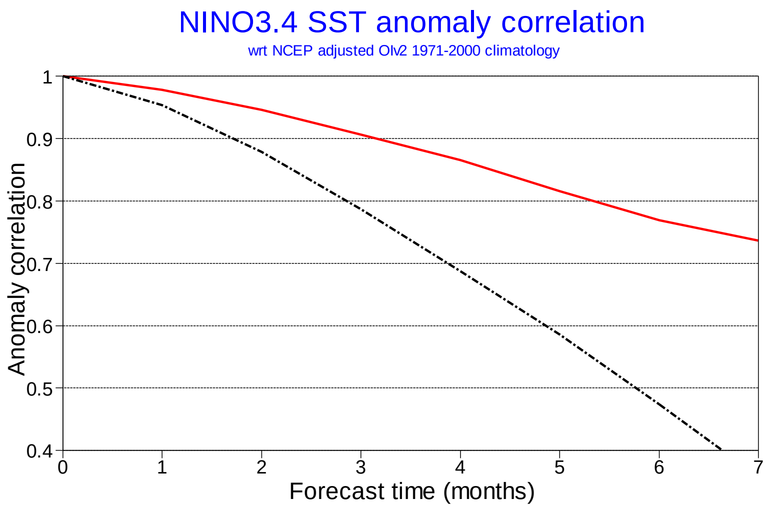
More recent ENSO forecasts are better ....

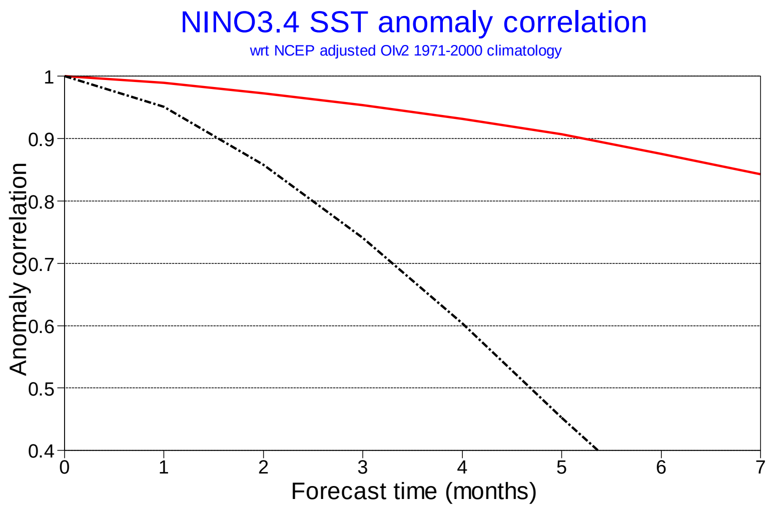




1981-1995

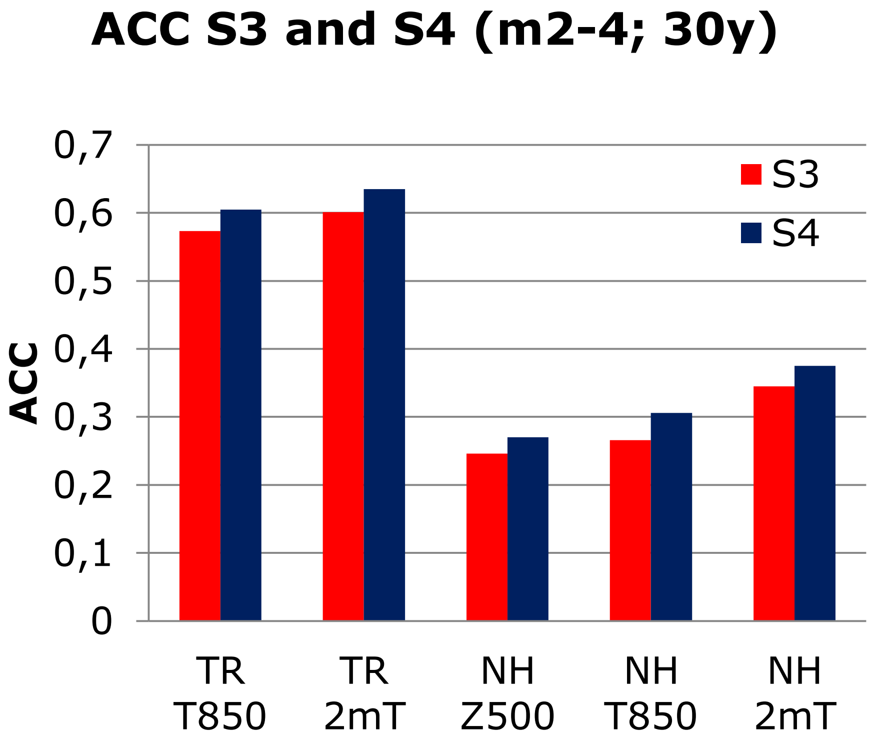
1996-2010



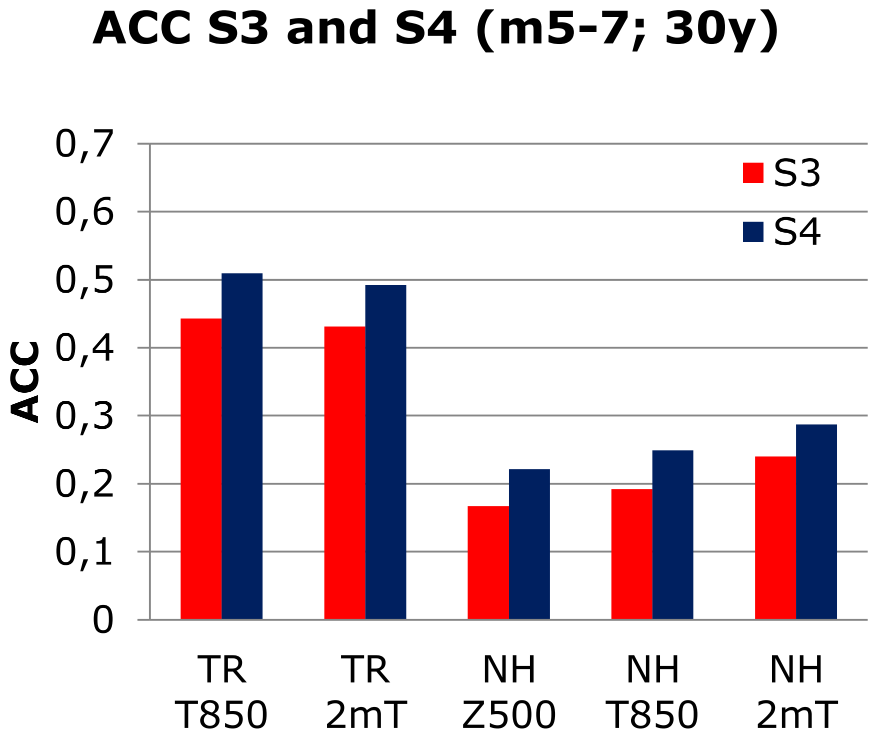


Tropospheric scores

Spatially averaged (with Fisher z-transform) grid-point temporal ACC
Scores for 1981-2010, aggregated over all 12 start months
NH is poleward of 30N, Tropics is 30N-30S



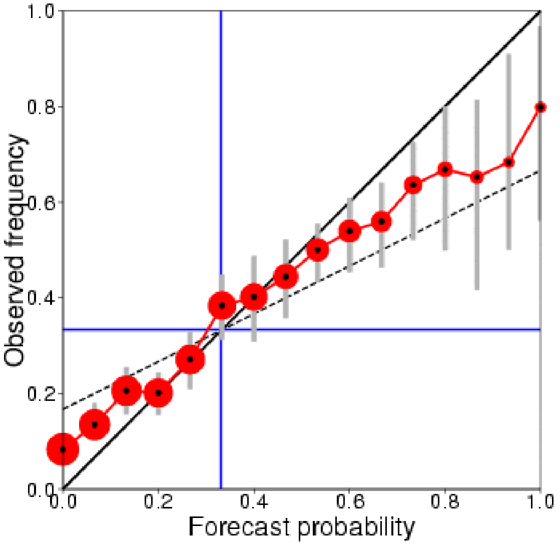
One month lead



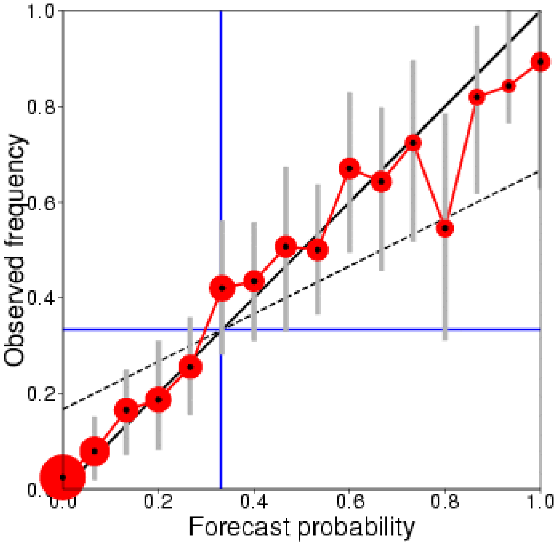
Four month lead

Probabilistic scores: Tropics

Reliability diagram for ECMWF with 15 ensemble members
Near-surface air temperature anomalies above the upper tercile
Accumulated over Africa (land points only)
Hindcast period 1981-2010 with start in May average over months 2 to 4
Skill scores and 95% conf. intervals (1000 samples)
Brier skill score: 0.129 (0.023, 0.202)
Reliability skill score: 0.975 (0.925, 0.988)
Resolution skill score: 0.154 (0.093, 0.219)



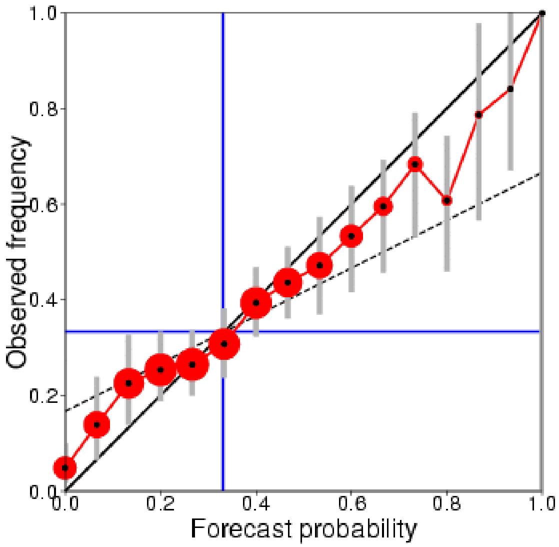
Reliability diagram for ECMWF with 15 ensemble members
Near-surface air temperature anomalies above the upper tercile
Accumulated over Southeast Asia (land points only)
Hindcast period 1981-2010 with start in May average over months 2 to 4
Skill scores and 95% conf. intervals (1000 samples)
Brier skill score: 0.328 (0.158, 0.451)
Reliability skill score: 0.982 (0.921, 0.987)
Resolution skill score: 0.346 (0.226, 0.474)



Probabilistic scores: Europe

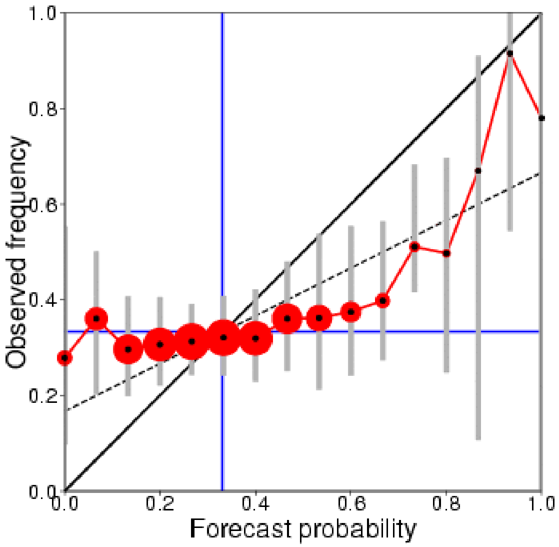
S4: JJA 2mT from 1st May

Reliability diagram for ECMWF with 15 ensemble members
Near-surface air temperature anomalies above the upper tercile
Accumulated over Europe (land and sea points)
Hindcast period 1981-2010 with start in May average over months 2 to 4
Skill scores and 95% conf. intervals (1000 samples)
Brier skill score: 0.092 (0.007, 0.162)
Reliability skill score: 0.986 (0.950, 0.994)
Resolution skill score: 0.106 (0.056, 0.173)



S4: DJF 2mT from 1st Nov

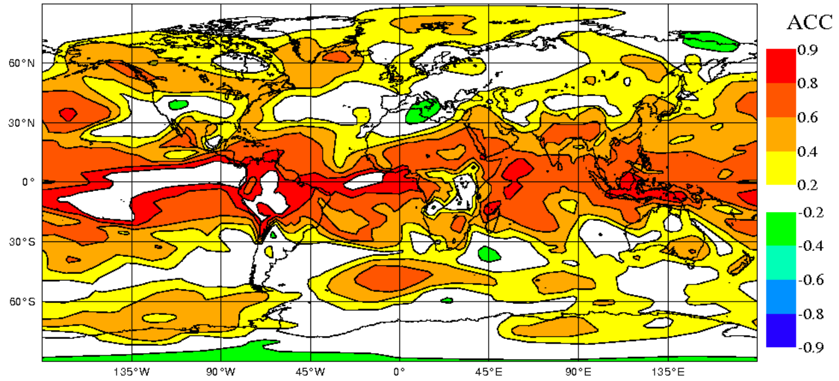
Reliability diagram for ECMWF with 15 ensemble members
Near-surface air temperature anomalies above the upper tercile
Accumulated over Europe (land and sea points)
Hindcast period 1981-2010 with start in November average over months 2 to 4
Skill scores and 95% conf. intervals (1000 samples)
Brier skill score: -0.081 (-0.191, 0.011)
Reliability skill score: 0.908 (0.790, 0.965)
Resolution skill score: 0.011 (0.006, 0.053)



S4 extended hindcast set

T850 Anom. correlation S4(15)-ERA Int 1981-2010DJF

Global z-mean acc: 0.483 NH:0.287 TR:0.644 SH:0.254

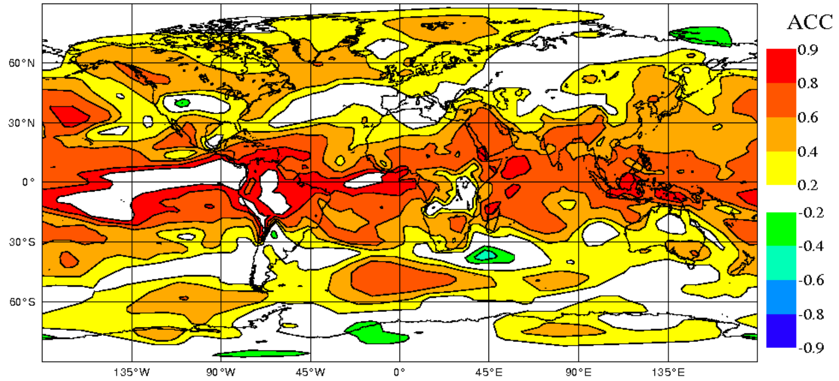


15 members

NH:0.287

T850 Anom. correlation S4(51)-ERA-Int 1981-2010DJF

Global z-mean acc: 0.505 NH:0.329 TR:0.658 SH:0.275

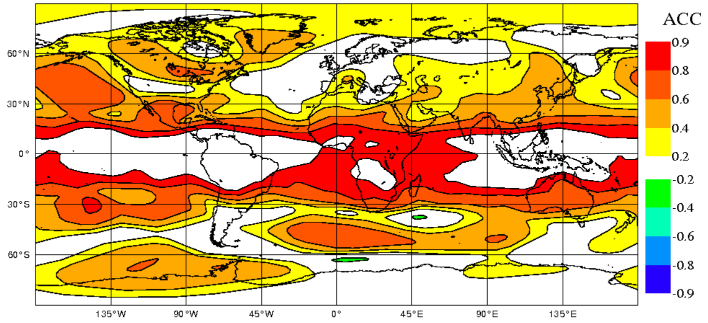


Scores are smoother and systematically higher with 51 member hindcasts

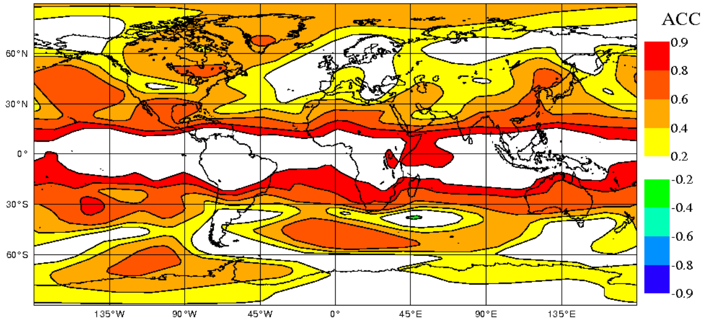
NH:0.329

S4 extended hindcast set

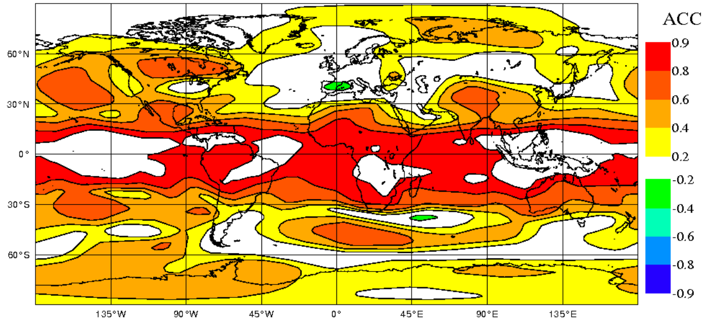
Z500 Anom. correlation S4(15)-ERA Int 1981-2010DJF
Global z-mean acc: 0.65 NH:0.331 TR:0.827 SH:0.355



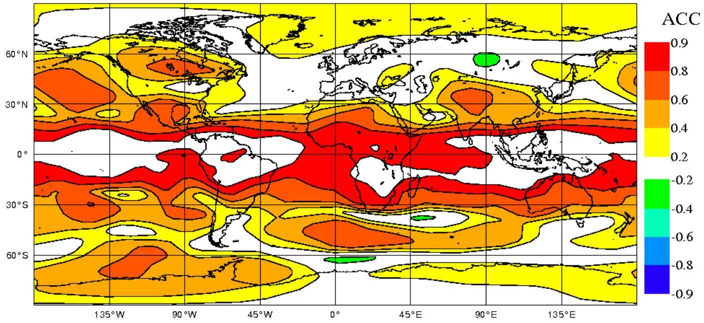
Z500 Anom. correlation S4(41)-ERA Int 1981-2010DJF
Global z-mean acc: 0.676 NH:0.381 TR:0.839 SH:0.397



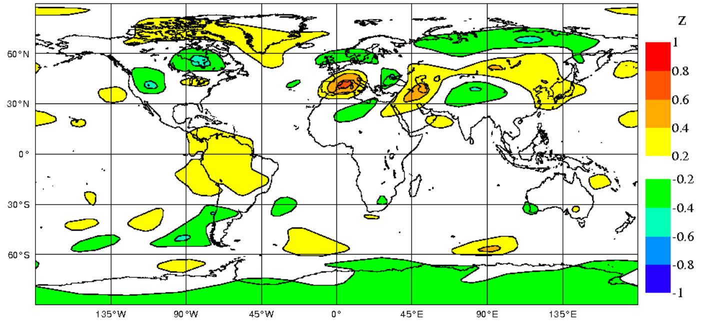
Z500 Anom. correlation S3(15)-ERA Int 1981-2010DJF
Global z-mean acc: 0.632 NH:0.301 TR:0.81 SH:0.373



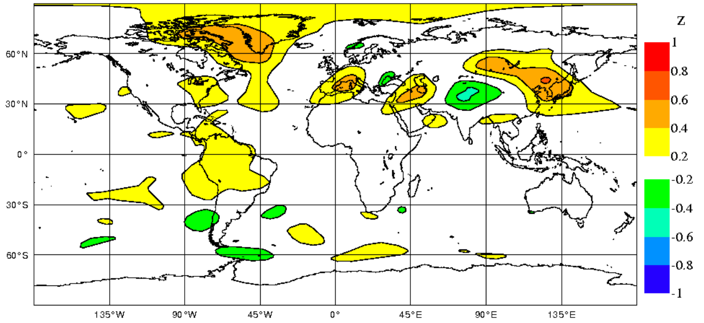
Z500 Anom. correlation S3(41)-ERA Int 1981-2010DJF
Global z-mean acc: 0.634 NH:0.277 TR:0.813 SH:0.388



Fisher z transform diff S4(15)-S3(15) 1981-2010DJF
sigma: 0.272 mean: 0.0303



Fisher z transform diff S4(41)-S3(41) 1981-2010DJF
sigma: 0.272 mean: 0.073



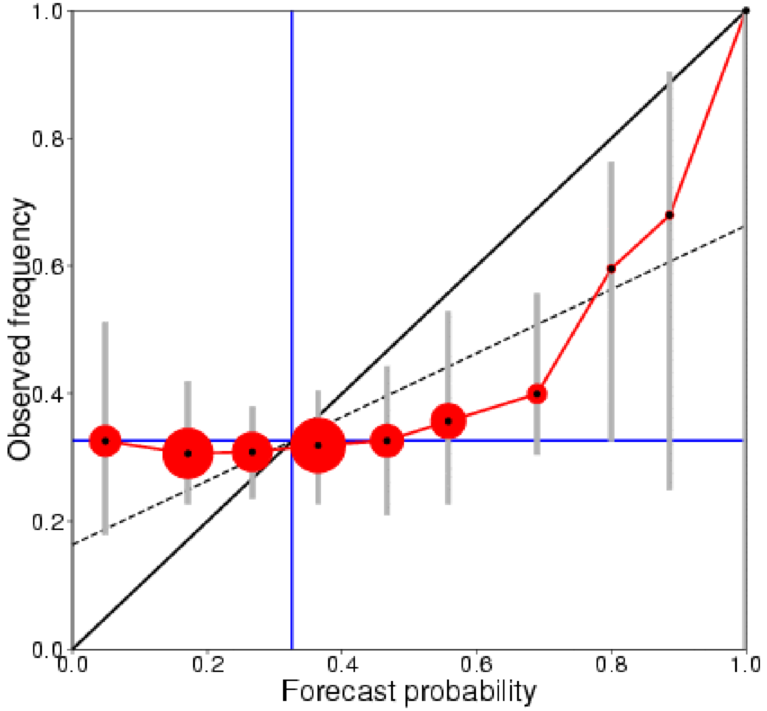
Gain over S3 is now stronger and more robust

(41 members each vs 15 members each)

S4 extended hindcast set

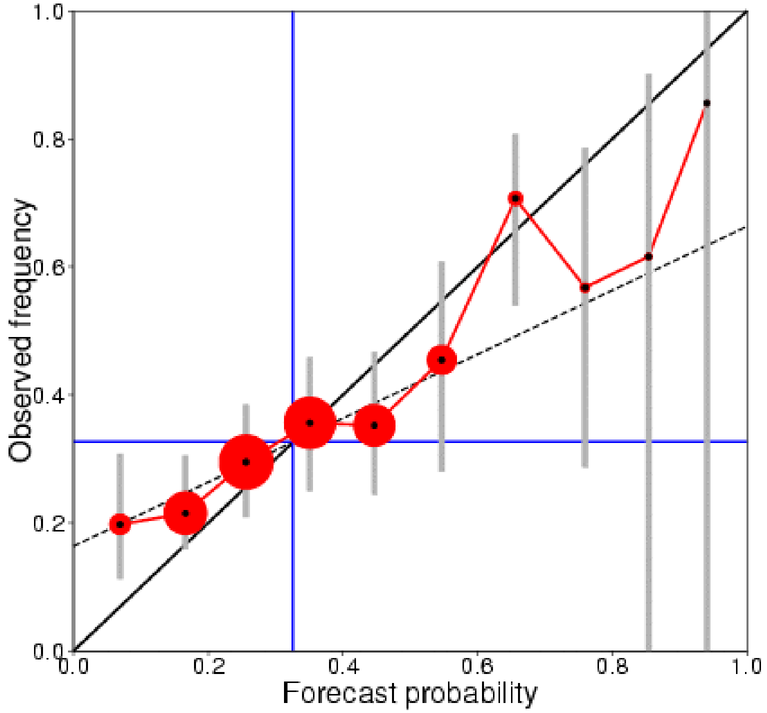
15 members

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

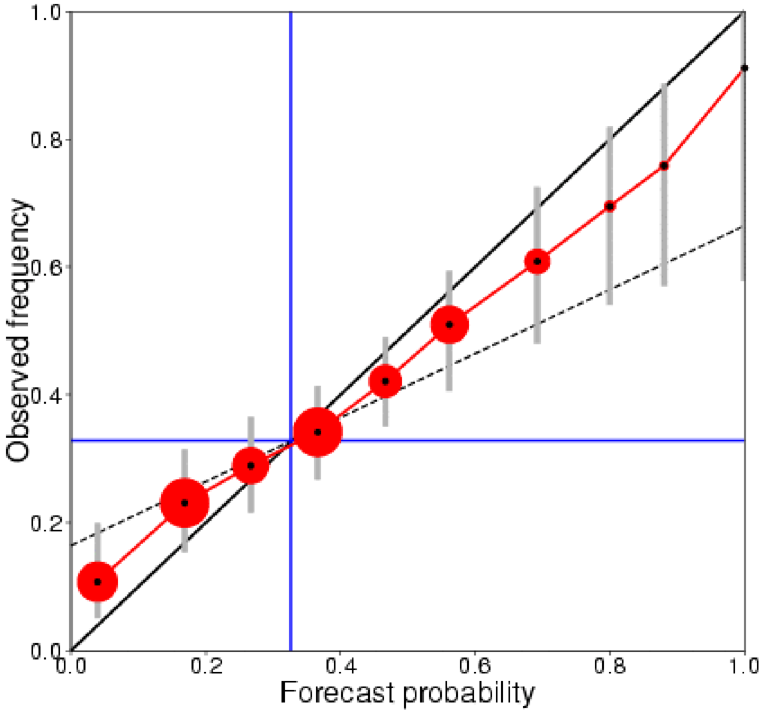


(Figures from Susanna Corti)

S4 extended hindcast set

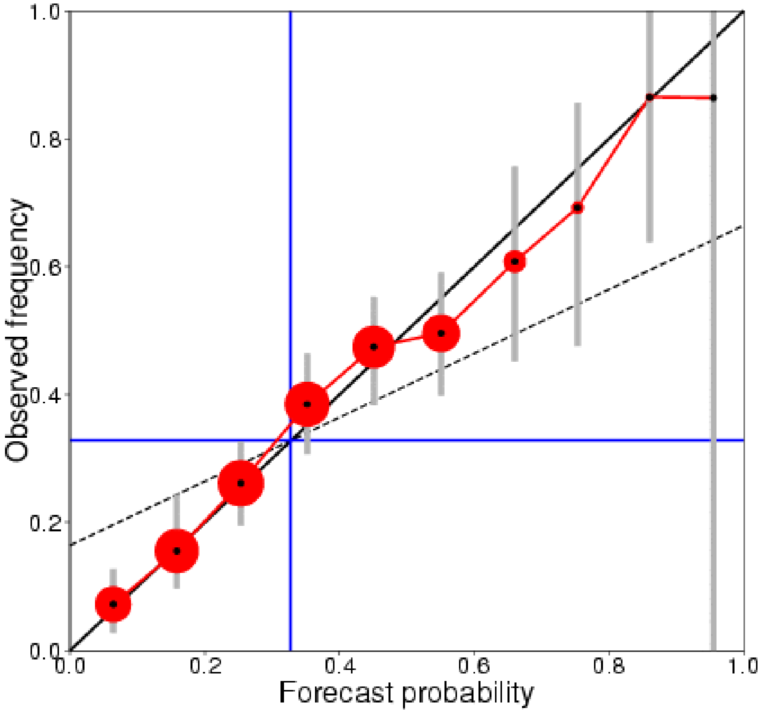
15 members

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



51 members

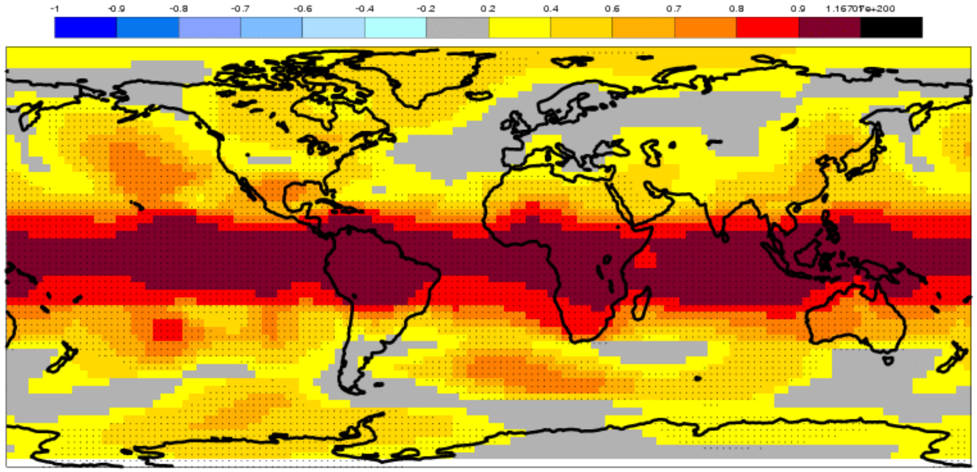
JJA Europe T2m>upper tercile
Re-forecasts from 1 May, 1981-2010
Reliability score: 0.996
ROC skill score: 0.43



(Figures from Susanna Corti)

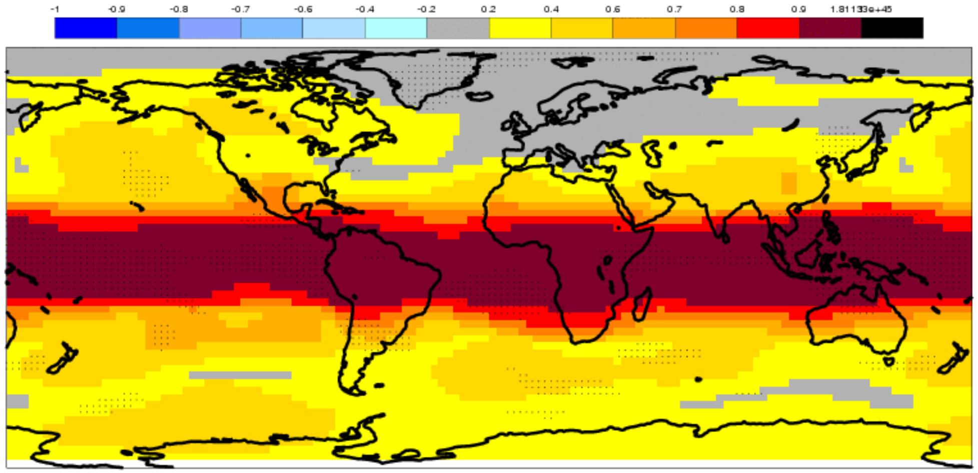
S4 ACC
DJF Z500

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)

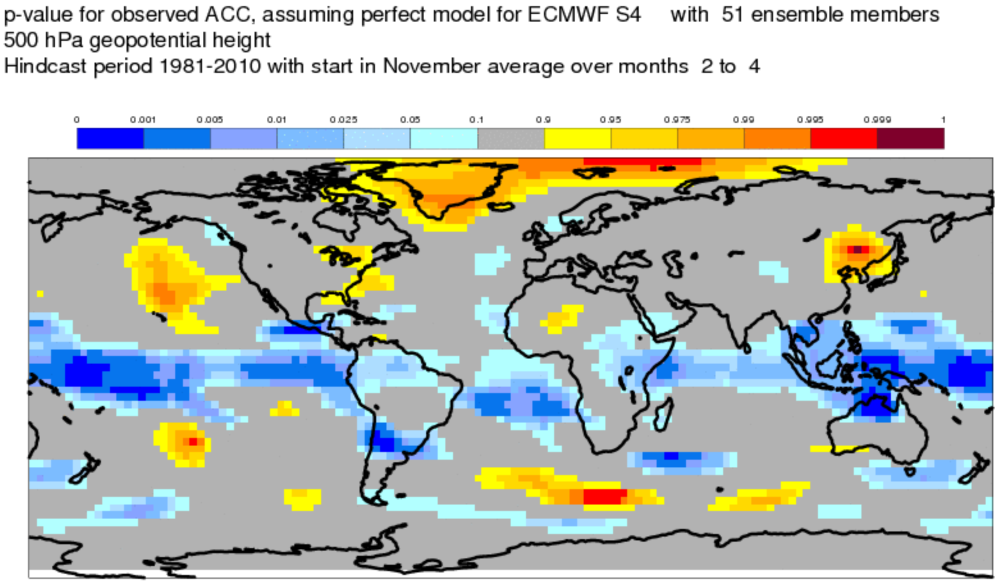


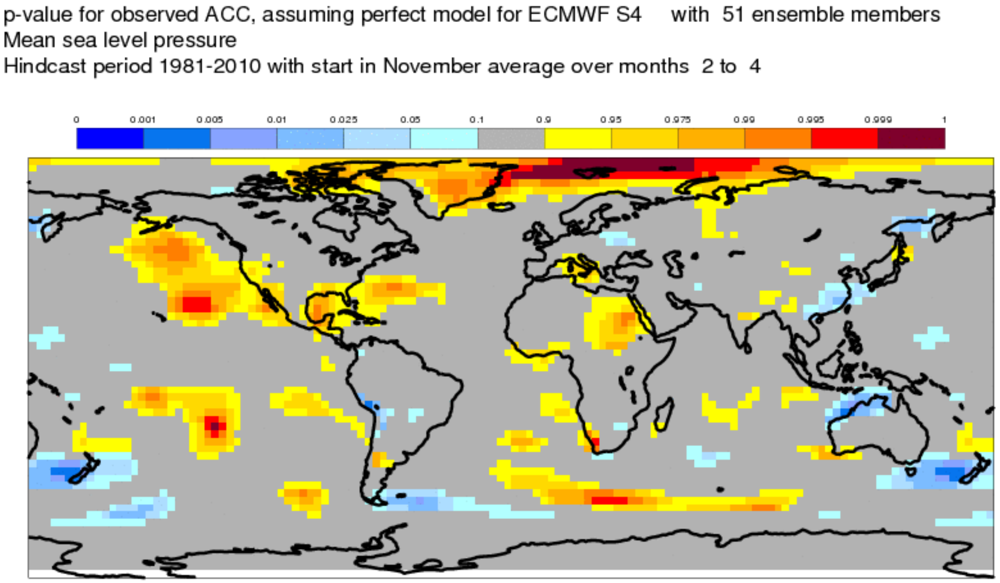
S4 ACC perfect model limit

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

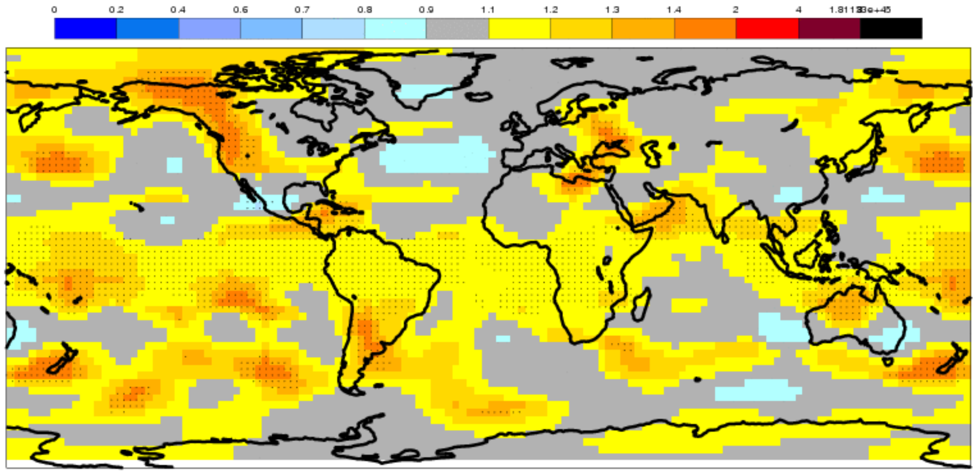

p-value for observed ACC, assuming perfect model for ECMWF S4 with 51 ensemble members
500 hPa geopotential height
Hindcast period 1981-2010 with start in November average over months 2 to 4


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

Model/observed variability

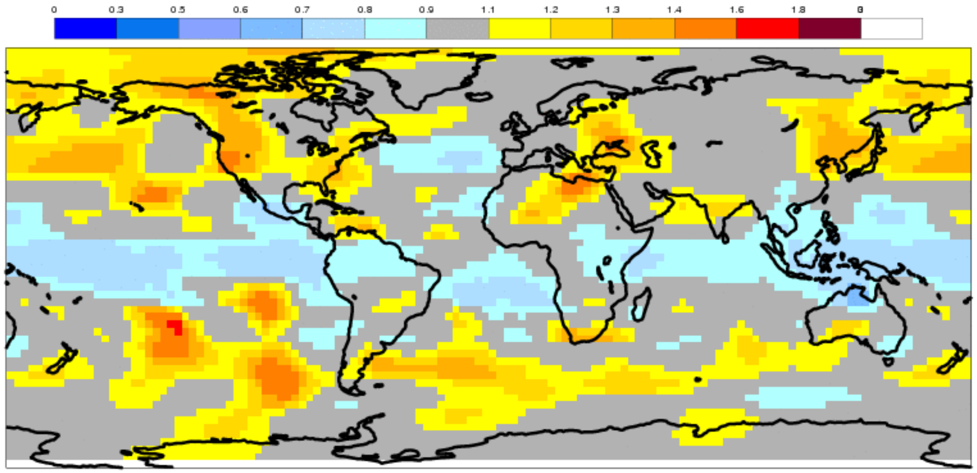
Ratio of SD (model/reference) for ECMWF S4 with 51 ensemble members
500 hPa geopotential height
Hindcast period 1981-2010 with start in November average over months 2 to 4
Black dots for values significantly different from zero with 95% confidence (1000 samples)



NH stddev ratio: 1.064
p val for observed stddev: 0.0785
NH stddev ratio 95% interval: 0.979 - 1.149

Ensemble spread / r.m.s. error

Ratio Spread(sd)/RMSE 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)



More significance testing

- **NH score (>30N), DJF Z500**
  - 30 years, 51 members: NH averaged ACC=0.358

- **What is the long-term average ACC?**
  - Bootstrap over nyears: 0.274 - 0.432

- **For these 30 years, what ACC would we get if model perfect?**
  - Expected value: 0.306
  - Bootstrap for a single ACC over internal sampling: 0.224 - 0.380
  - p-value for actual ACC: 0.088

- **For these 30 years, what is the sampling error for nens=51?**
  - Jackknife estimate for nens=inf: 0.384
  - Jackknife 95% interval: 0.335 – 0.431

Testing model versions

T159 expts, proposed new cycle

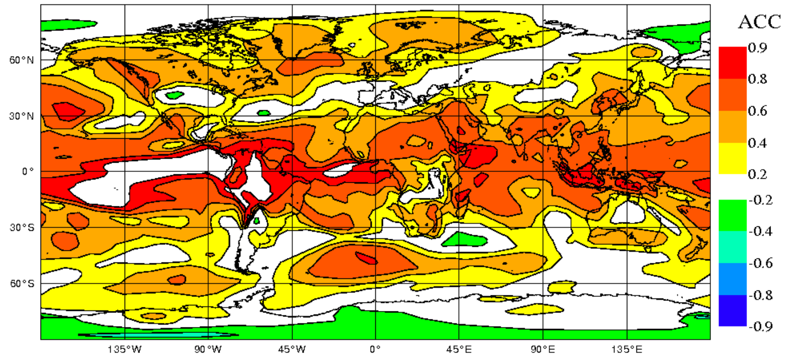
fuhg: vertical diffusion change
fulf : control

30 years, 101 members each

Skill difference is very large – and is significant with this sample size

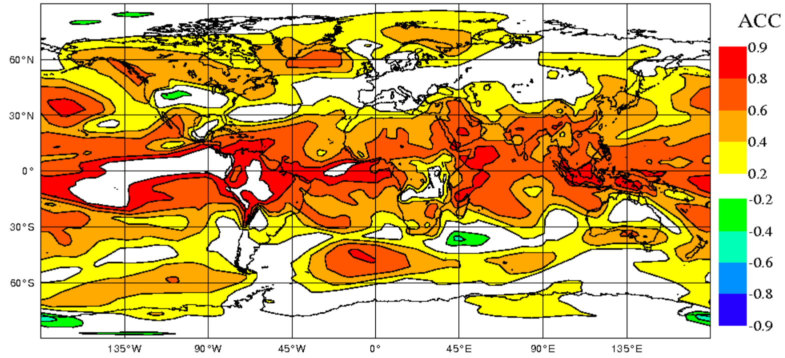
T850 Anom. correlation fuhg(101)-ERA-Int 1981-2010DJF

Global z-mean acc: 0.513 NH:0.351 TR:0.662 SH:0.28



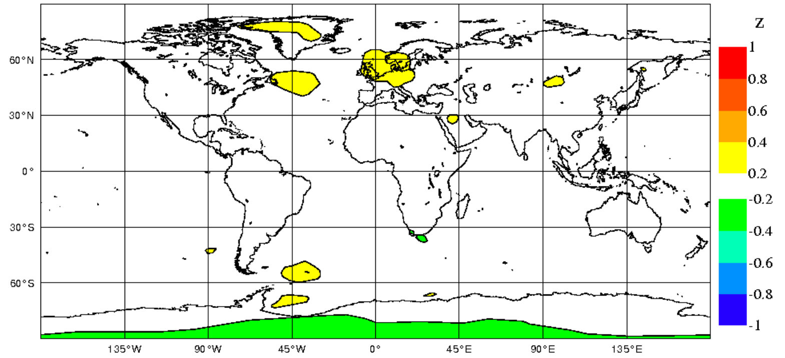
T850 Anom. correlation fulf(101)-ERA-Int 1981-2010DJF

Global z-mean acc: 0.503 NH:0.309 TR:0.66 SH:0.283



Fisher z transform diff fuhg(101)-fulf(101) 1981-2010DJF

sigma: 0.272 mean: 0.0133

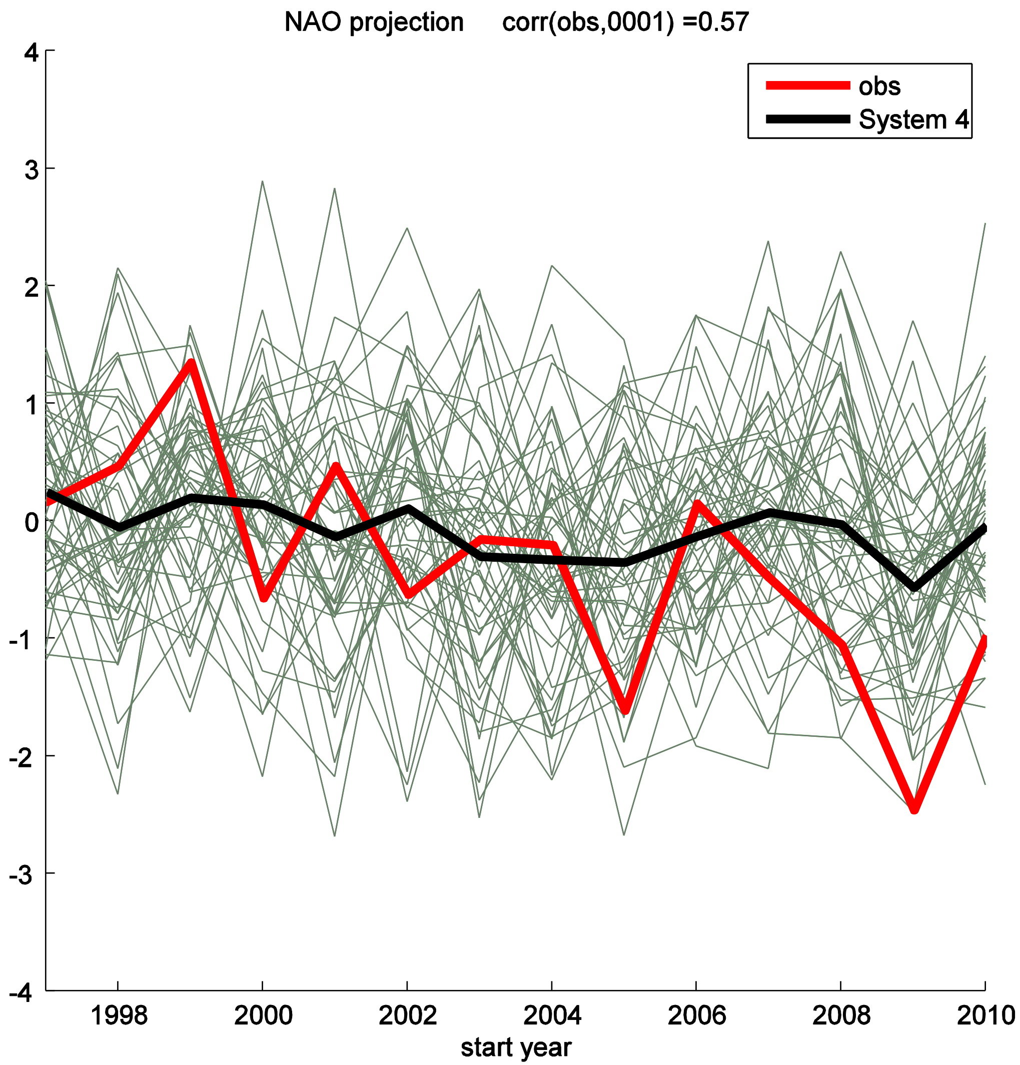


NAO statistics

Expt	Period	Ens. size	NAO acc
System 4	1981-2010	15	0.24
System 4	1981-2010	51	0.38
System 4	1997-2010	51	**0.57**
System 3	1981-2010	41	0.25
System 3	1997-2010	41	0.30

(PNA=0.70)

(NAO by projection onto observed Z500 pattern)



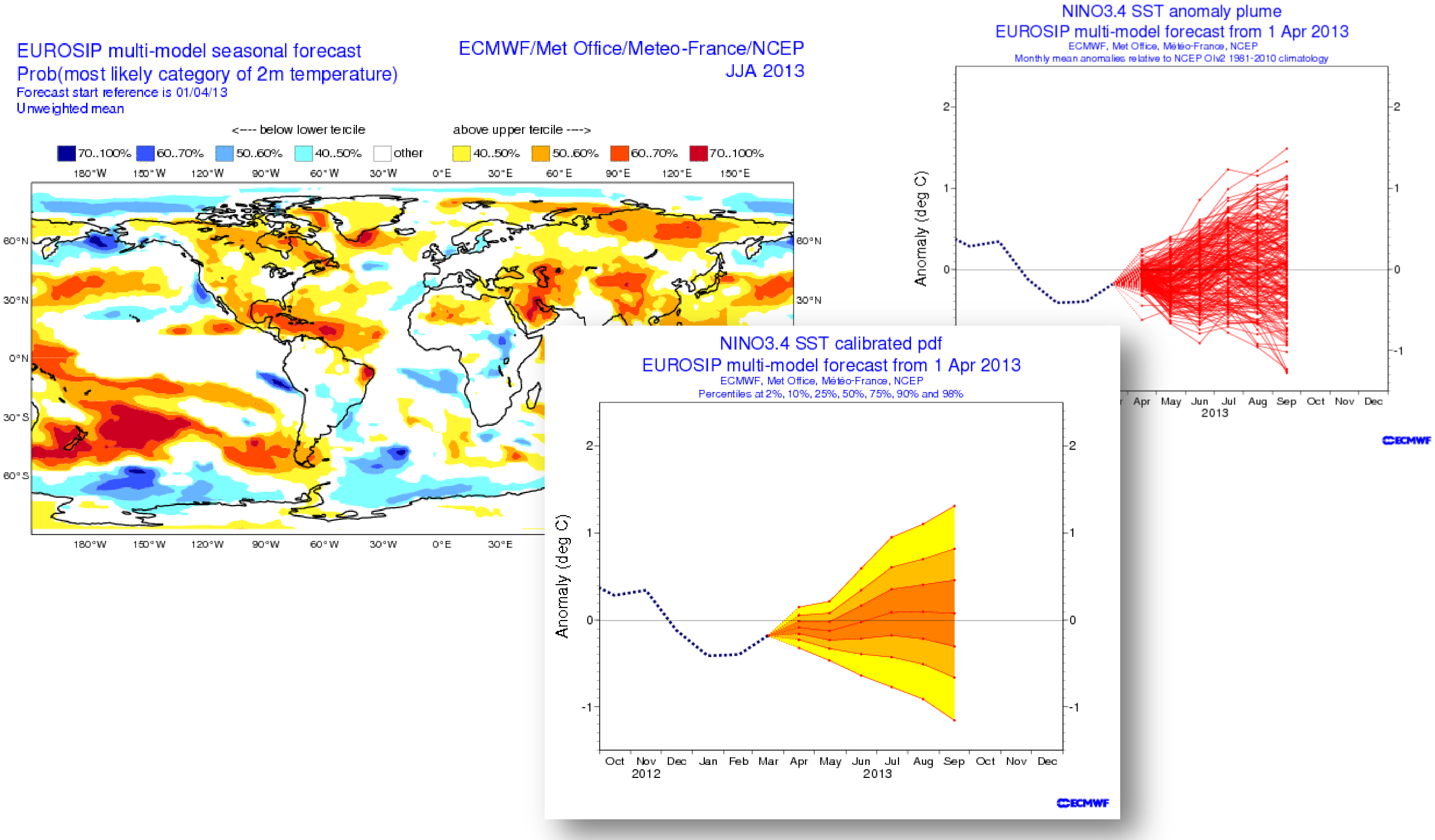
(Figure from Antje Weisheimer)

EUROSIP

- **A European multi-model seasonal forecast system**
  - Reliable, operational real-time system
  - Data archive, especially for research
  - Real-time forecast products
  - Operational from 2005
- **Implementation**
  - Partners: ECMWF, Met Office, Météo-France
  - Associate partner: NCEP

  - Expected future partners: DWD and possibly others
- **Regional approach, c.f. NMME, APCC**

EUROSIP web products



EUROSIP multi-model seasonal forecast
Prob(most likely category of 2m temperature)
Forecast start reference is 01/04/13
Unweighted mean

ECMWF/Met Office/Meteo-France/NCEP
JJA 2013

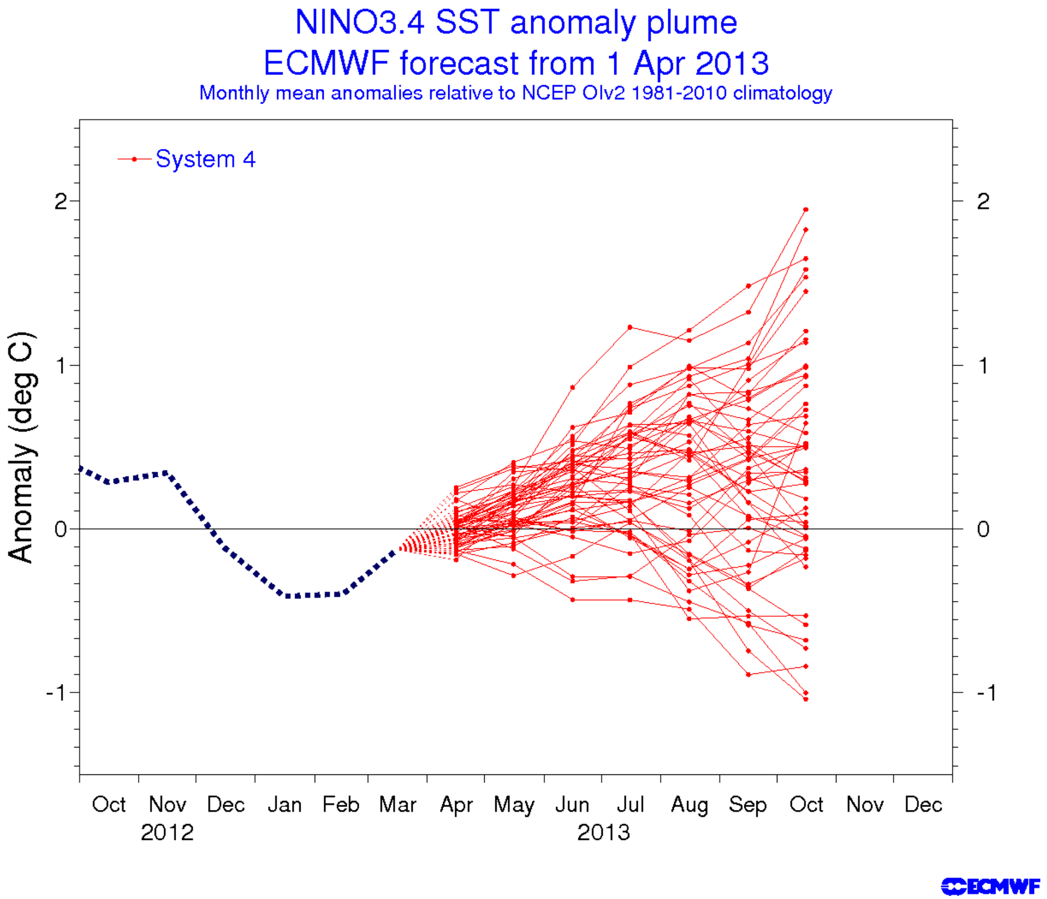
NINO3.4 SST anomaly plume
EUROSIP multi-model forecast from 1 Apr 2013
ECMWF, Met Office, Météo-France, NCEP
Monthly mean anomalies relative to NCEP OIv2 1981-2010 climatology

NINO3.4 SST calibrated pdf
EUROSIP multi-model forecast from 1 Apr 2013
ECMWF, Met Office, Météo-France, NCEP
Percentiles at 2%, 10%, 25%, 50%, 75%, 90% and 98%

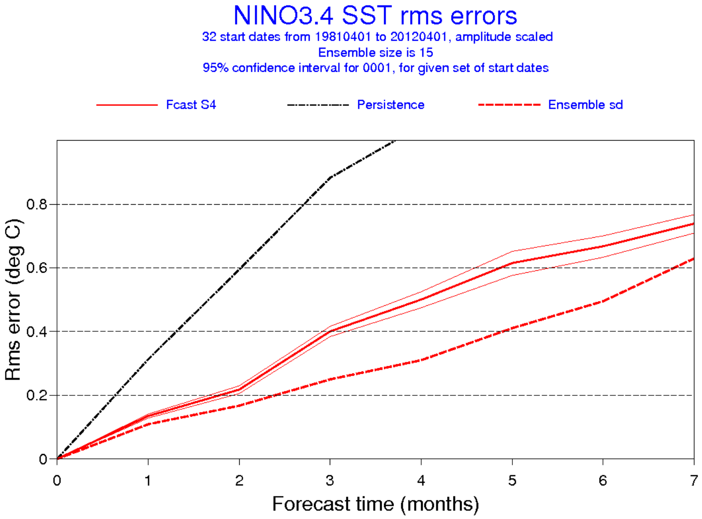
NINO SST pdf estimation

- **Parameterized, calibrated fit**
  - t-distribution, allowing for uncertainties in skill estimate
  - Calibrated against past performance
  - Rank histograms verify well
- **Robust implementation**
  - Weighted with past skill, but very conservatively
  - Predicted uncertainty only partially dependent on inter-model spread
- **pdf interpretation**
  - Based on past errors, doesn’t account for extreme risks
  - Bayesian probabilities: other systems will give a different answer, but both are correct

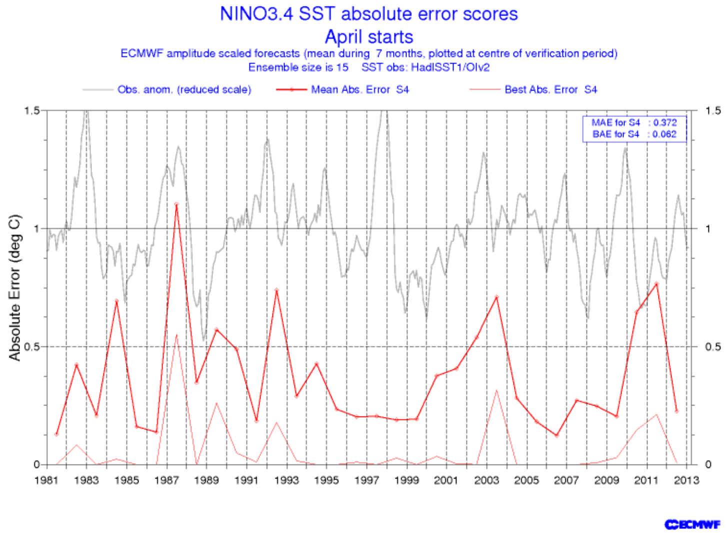
ECMWF forecast: ENSO



NINO3.4 SST anomaly plume
ECMWF forecast from 1 Apr 2013
Monthly mean anomalies relative to NCEP OIv2 1981-2010 climatology



NINO3.4 SST rms errors
32 start dates from 19810401 to 20120401, amplitude scaled
Ensemble size is 15
95% confidence interval for 0001, for given set of start dates

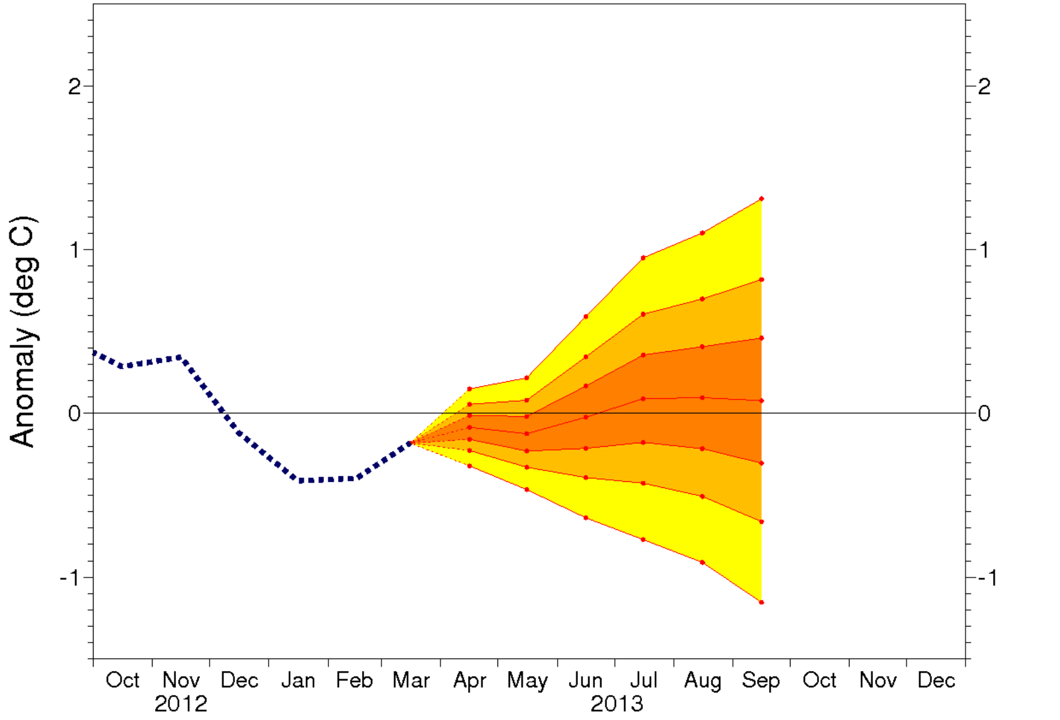


NINO3.4 SST absolute error scores
April starts
ECMWF amplitude scaled forecasts (mean during 7 months, plotted at centre of verification period)
Ensemble size is 15 SST obs: HadISST1/OIv2

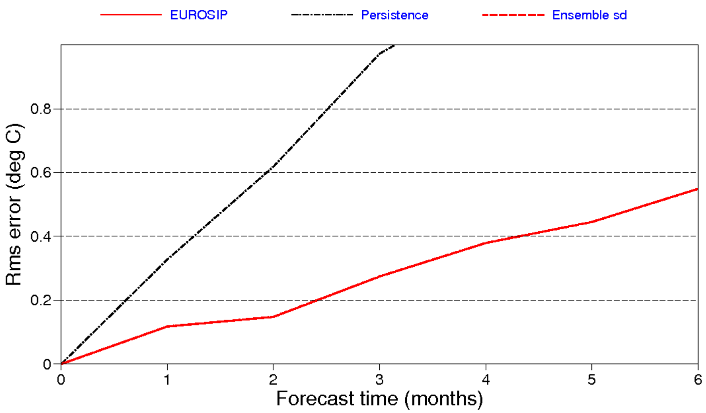
Past performance

EUROSIP forecast: ENSO

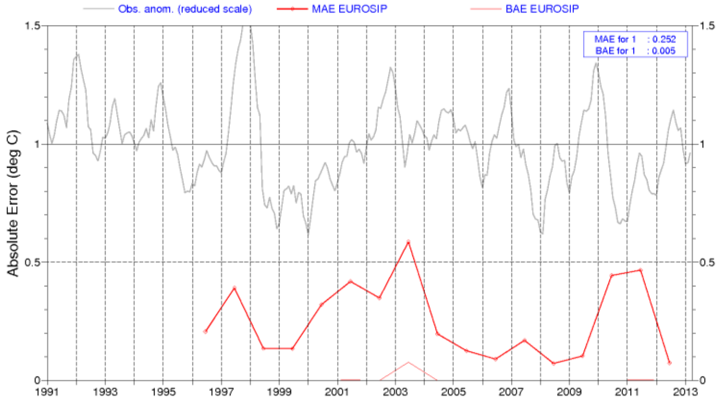
NINO3.4 SST calibrated pdf
EUROSIP multi-model forecast from 1 Apr 2013
ECMWF, Met Office, Météo-France, NCEP
Percentiles at 2%, 10%, 25%, 50%, 75%, 90% and 98%



NINO3.4 SST rms errors
17 start dates from 19960401 to 20120401, calibrated
Ensemble size is variable



NINO3.4 SST absolute error scores
April starts
Calibrated pdf forecasts (mean score over 6 months, plotted at centre of verification period)
Ensemble size is 0 SST obs: HadISST1/OIv2



Past performance

To conclude

- ECMWF S4 has a very good overall performance
- With 51 members, mid-latitudes look better than with 15
- NH winter – skill is better than expected given the model S/N ratio
- Implies predictability limit higher than model estimate

- Mid-latitude skill estimates are subject to large uncertainties, due to both ensemble size and number of years
- Need careful and appropriate tests and error bars. Don’t be too quick to draw conclusions, negative or positive. Small samples are often all we have.

- Multi-model forecasting is valuable, both for operations and research
- Scope remains for better calibrated products

- Exciting times …. .