The EUROSIP system

Tim Stockdale

Thanks to

Met Office, Météo-France, NCEP

Laura Ferranti, Frederic Vitart, Gianpaolo Balsamo

Outline

- **The multi-model concept**

  ~ Interlude ~

- **EUROSIP – past, present, future**
- **Calibrated forecasts**
- **Conclusions**

Model error

- **By model error we mean problems, inadequacies and imperfections with the model formulation and its numerical implementation.**

- **This model error causes integrations of the model to produce results which are unrealistic in various ways; e.g. the model climate (mean, variability, features) may be unrealistic.**

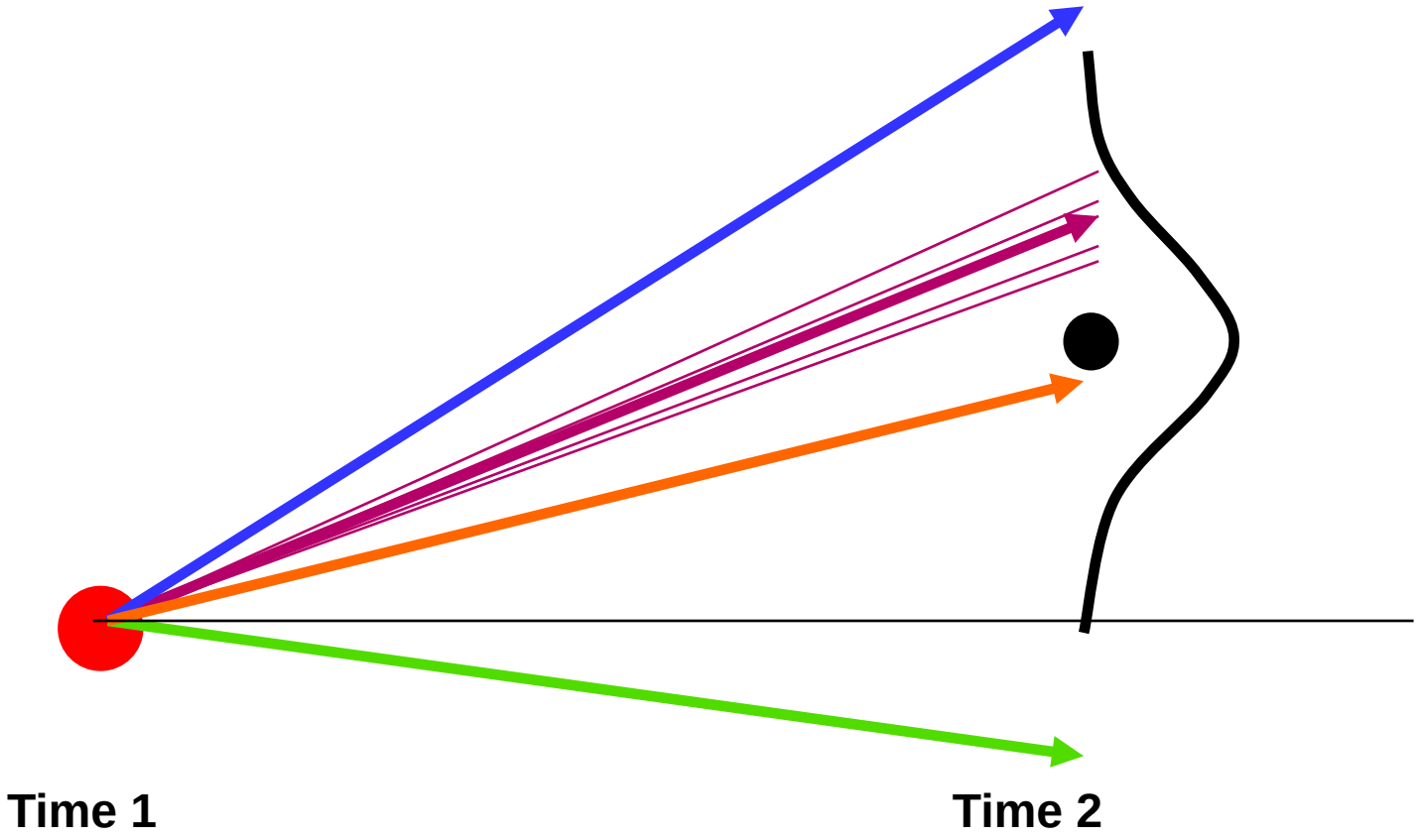
- **The imperfections in the model also contribute to errors in any seasonal forecast produced by the model. This contribution we define as the model forecast error. We do not know its value in any particular case, but may try to estimate its statistical properties.**

Multi-model ensemble

- **Different coupled GCMs have different model errors**
  - There may be lots of common errors, too.
- **So let's take an 'ensemble' of model forecasts:**
  - The mean of the ensemble should be better, because at least some of the model forecast errors will be averaged out
  - The 'spread' of the ensemble should be better, since we are sampling some of the uncertainty
- **An ensemble of *forecast values* or of *models?***

Multi-model ensemble of forecast values

- **What would an ‘ideal’ multi-model system look like?**
  - Assume fairly large number of models (10 or more)
  - Assume models have roughly equal levels of forecast error
  - Assume that model forecast errors are *uncorrelated*
  - Assume that each model has its own mean bias removed



Error in ensemble mean = σ / √n

Multi-model ensemble of forecast values

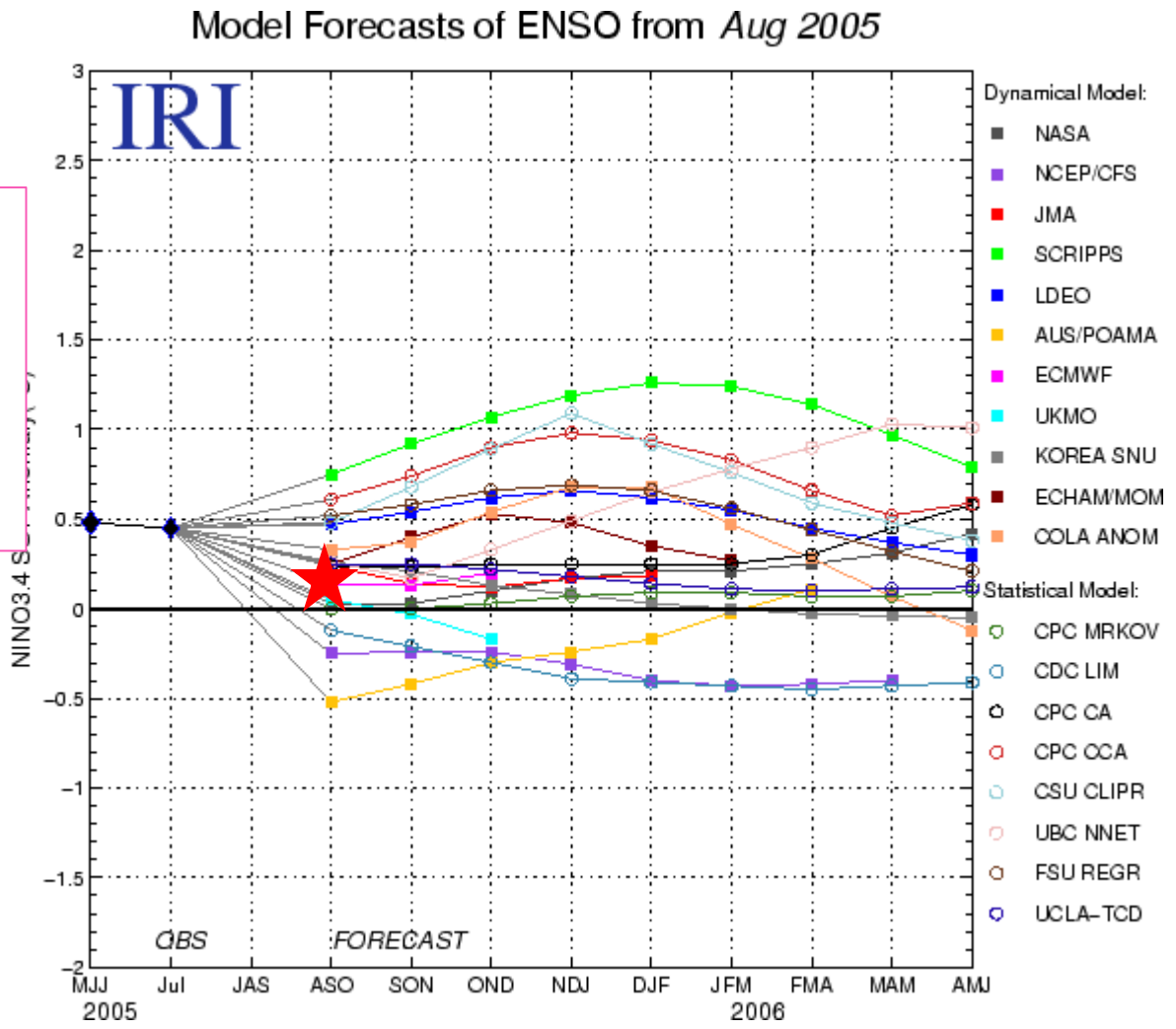
- **What would an 'ideal' multi-model system look like?**
  - Assume fairly large number of models (10 or more)
  - Assume models have roughly equal levels of forecast error
  - Assume that model forecast errors are *uncorrelated*
  - Assume that each model has its own mean bias removed
  - A priori, for each forecast, we consider each of the models' forecasts equally likely [in a Bayesian sense – in reality, all the model pdfs will be wrong]
  - A posteriori, this is no longer the case: forecasts near the centre of the multi-model distribution have higher likelihood
  - *Different* from a single model ensemble with perturbed ic's.
  - Multi-model ensemble distribution is NOT a pdf

Non-ideal case

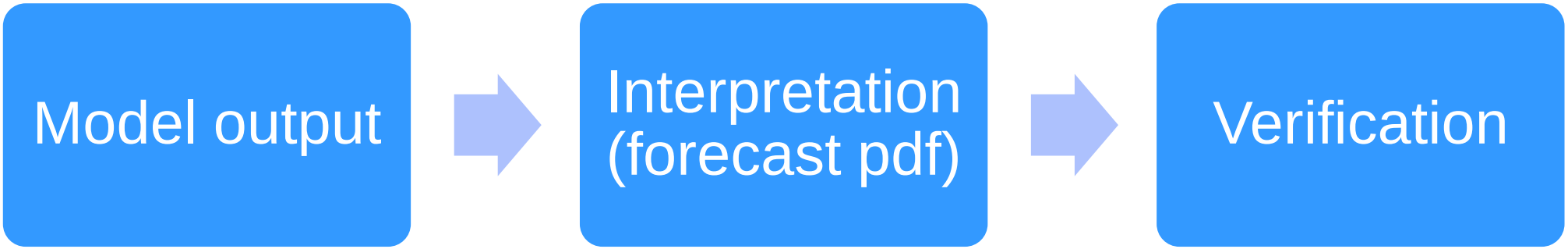
- **Model forecast errors are *not* independent**
  - Dependence will reduce degrees of freedom, hence the effective *n*; will increase uncertainty
  - In some cases, reduction in *n* could be drastic
- **Initial condition error can be important**
  - The foregoing analysis applies to the ‘model error’ contribution to error variance
  - Initial condition error and irreducible error growth terms follow usual ensemble behaviour, and must be accounted for separately
- **What weight should be given to outliers?**

Multi-model ensemble is not a pdf

Although we can choose to treat it as one if we want (and many people do).



Forecast process



Forecast pdf *should* be an appropriate interpretation of model ensemble, not an equivalence.

The DEMETER project (c.f. PROVOST and ENSEMBLES)

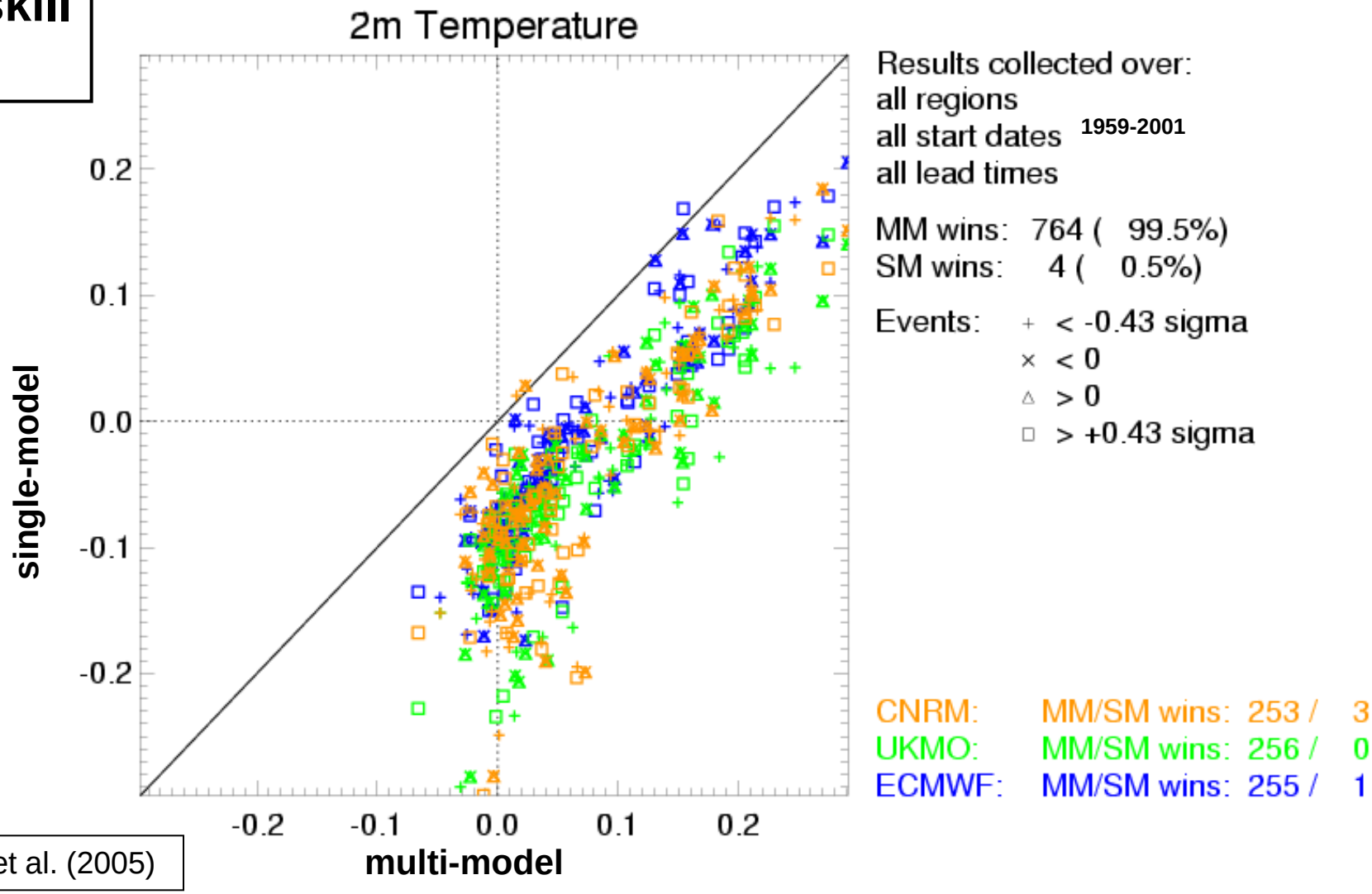
EU funded, 2000-2003
Multi-model study with 7 coupled general circulation models

Partner	Atmosphere	Ocean
ECMWF	IFS	HOPE
LODYC	IFS	OPA 8.3
CNRM	ARPEGE	OPA 8.1
CERFACS	ARPEGE	OPA 8.3
INGV	ECHAM-4	OPA 8.2
MPI	ECHAM-5	MPI-OM1
UKMO	HadCM3	HadCM3

http://www.ecmwf.int/research/demeter/

DEMETER: Brier score of multi-model vs single-model

Brier skill score

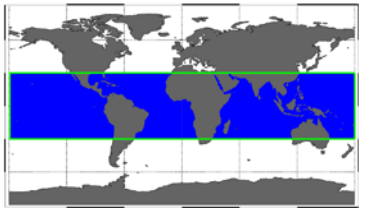


Hagedorn et al. (2005)

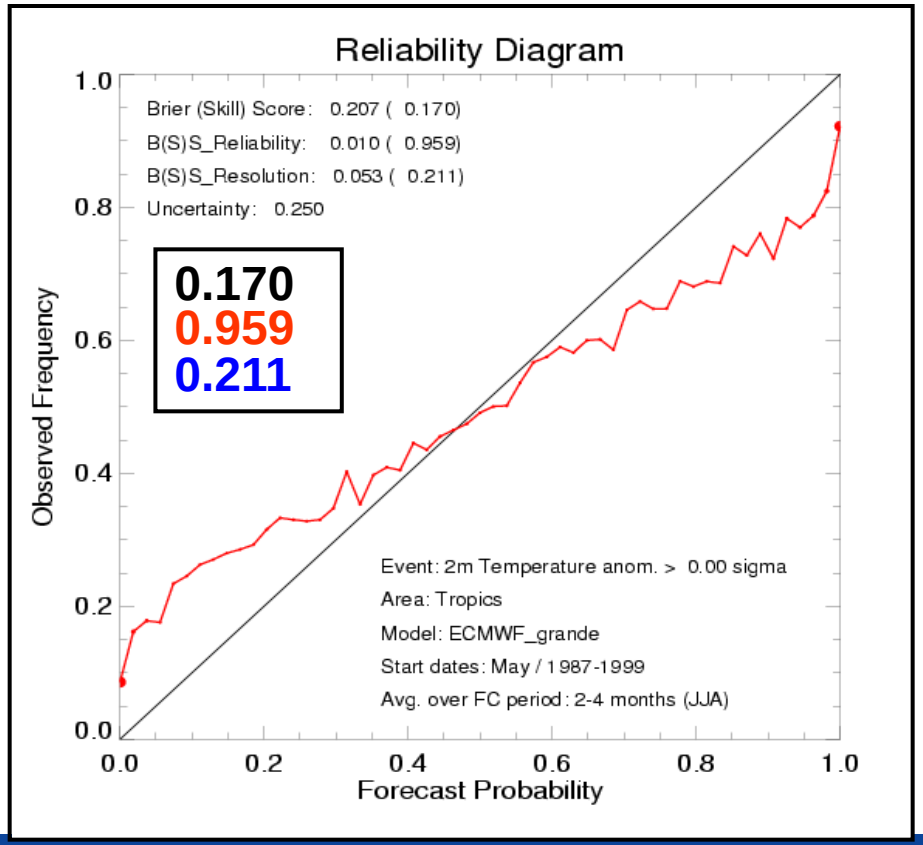
DEMETER: not just ensemble size

BSS
Rel-Sc
Res-Sc

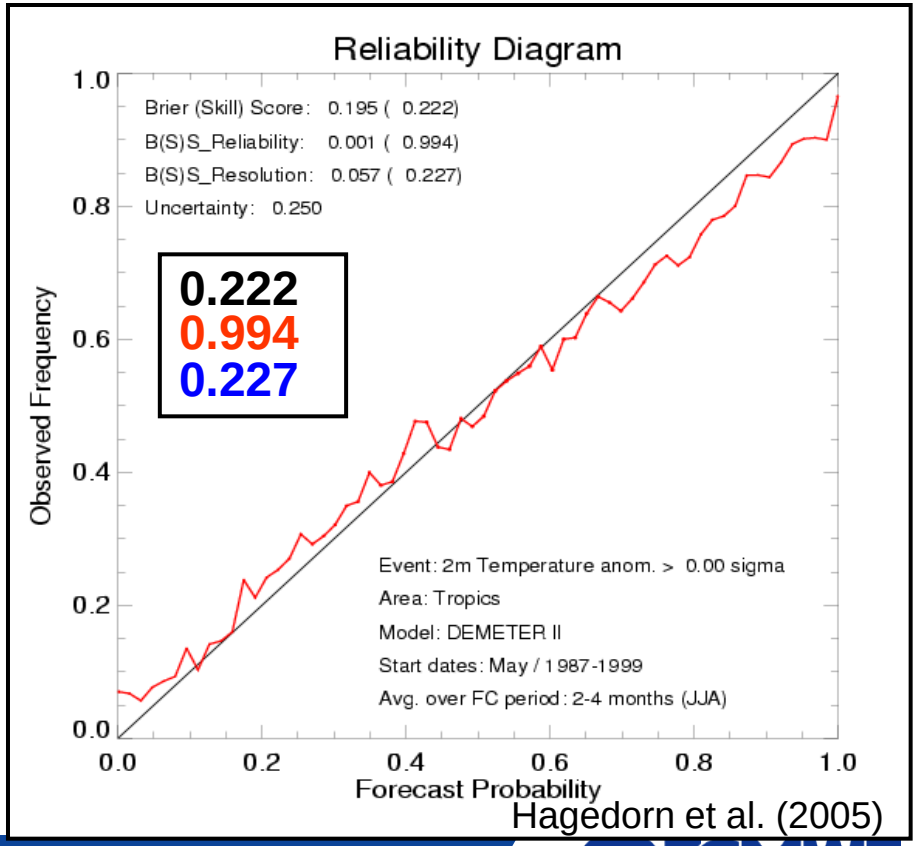
Reliability diagrams (T2m > 0)
1-month lead, start date May, 1987 - 1999



single-model [54 members]

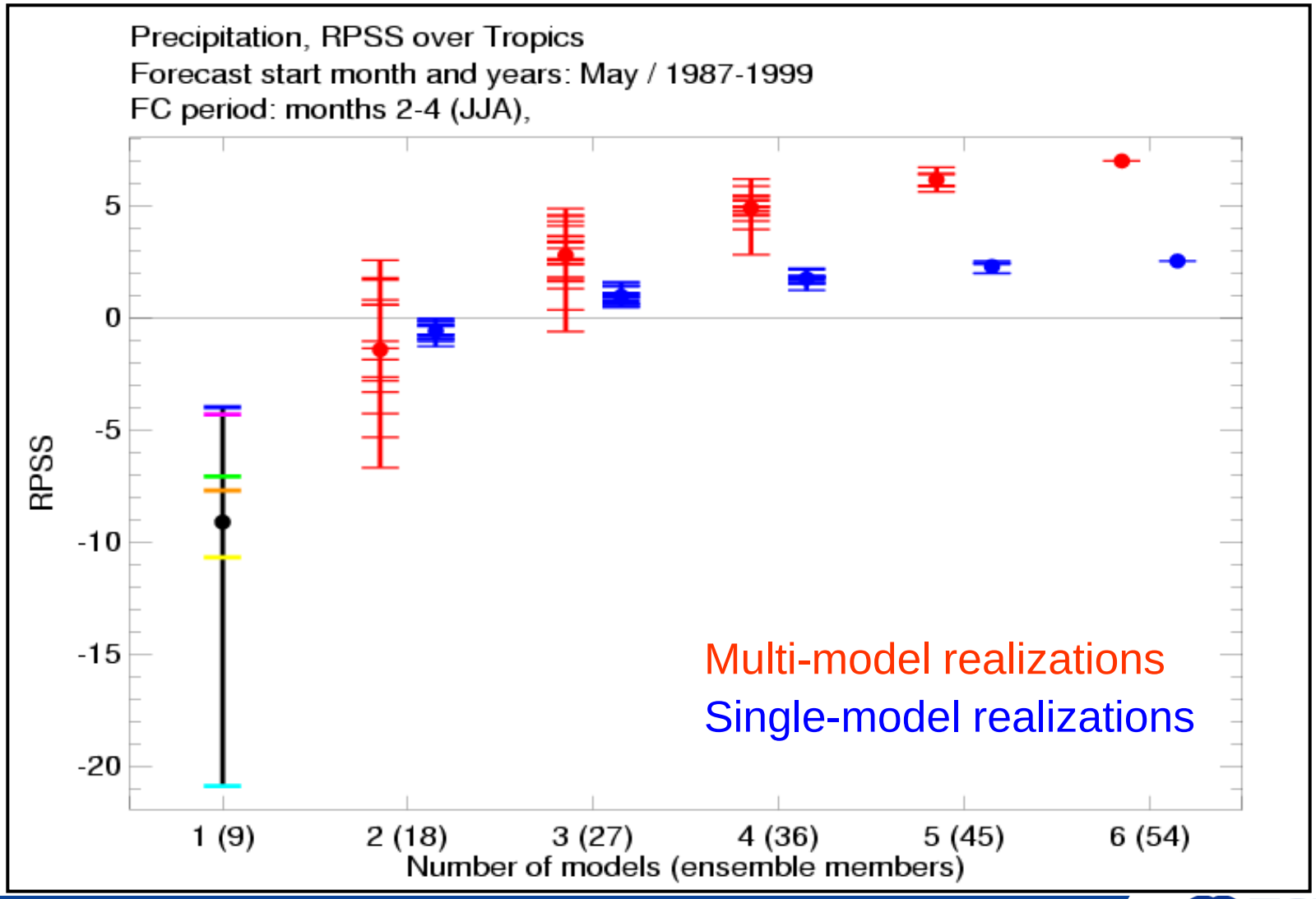


multi-model [54 members]



Hagedorn et al. (2005)

DEMETER: impact of number of models



~

Interlude

~

Additional comments on System 4

More recent ENSO forecasts are better ....

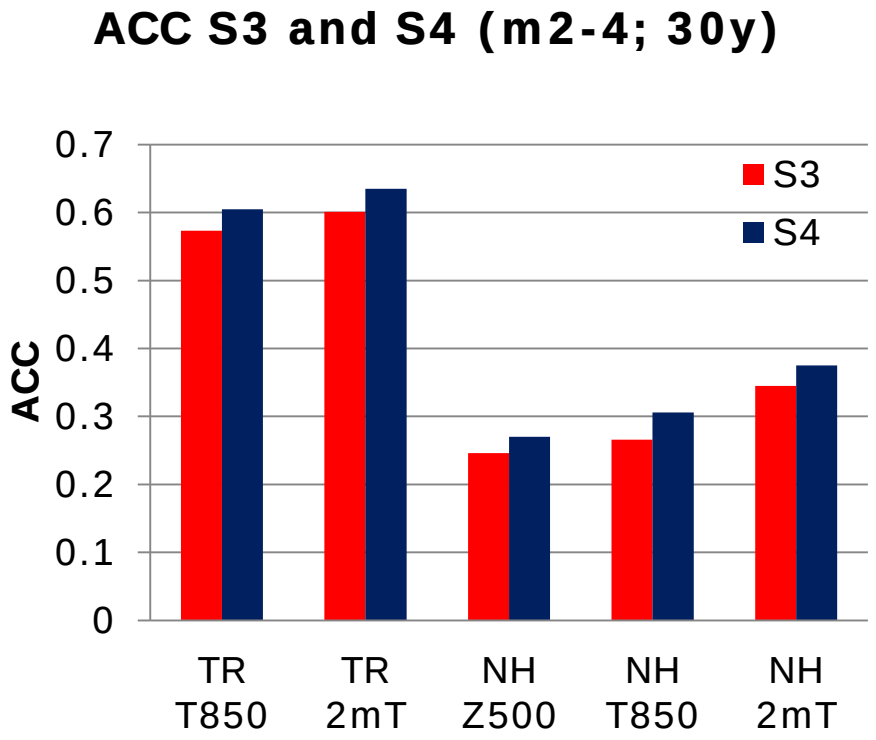
1981-1995

1996-2010

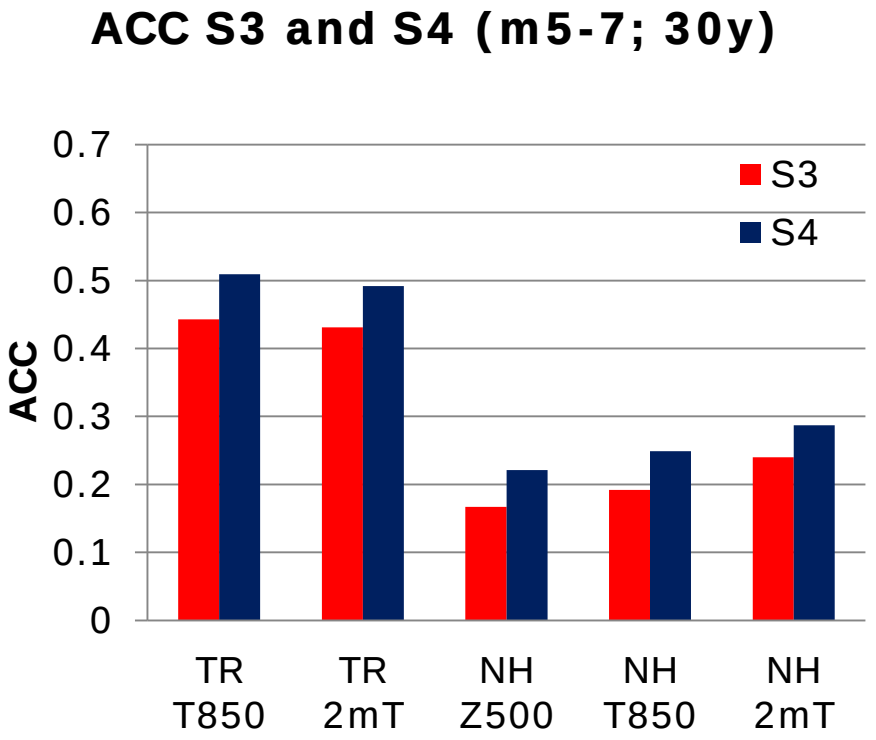
Reduced mean state errors

Tropospheric scores

Spatially averaged grid-point temporal ACC



One month lead



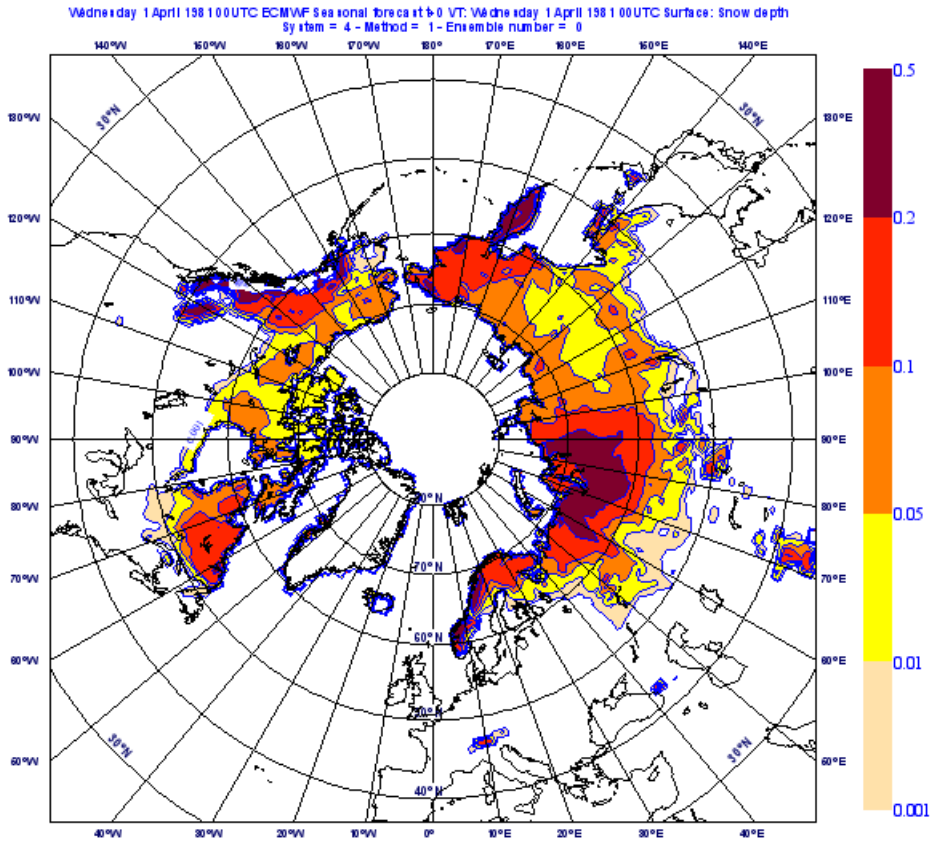
Four month lead

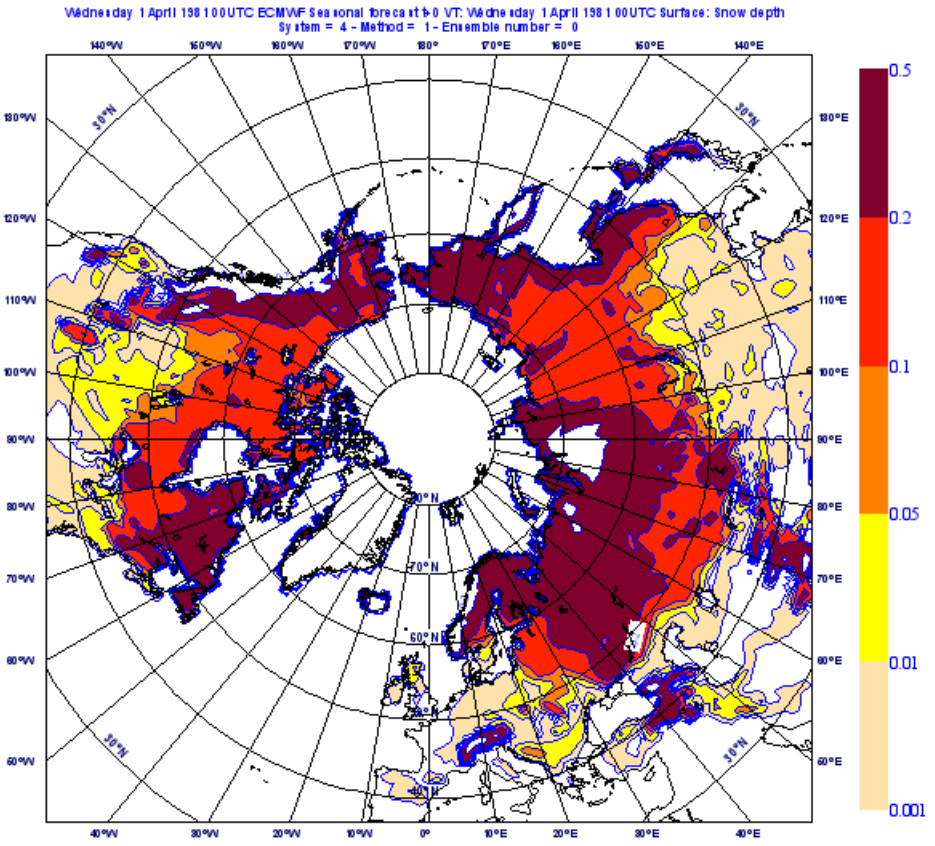
QBO

Problematic ozone analyses

Stratosphere improved in S4, but still problematic

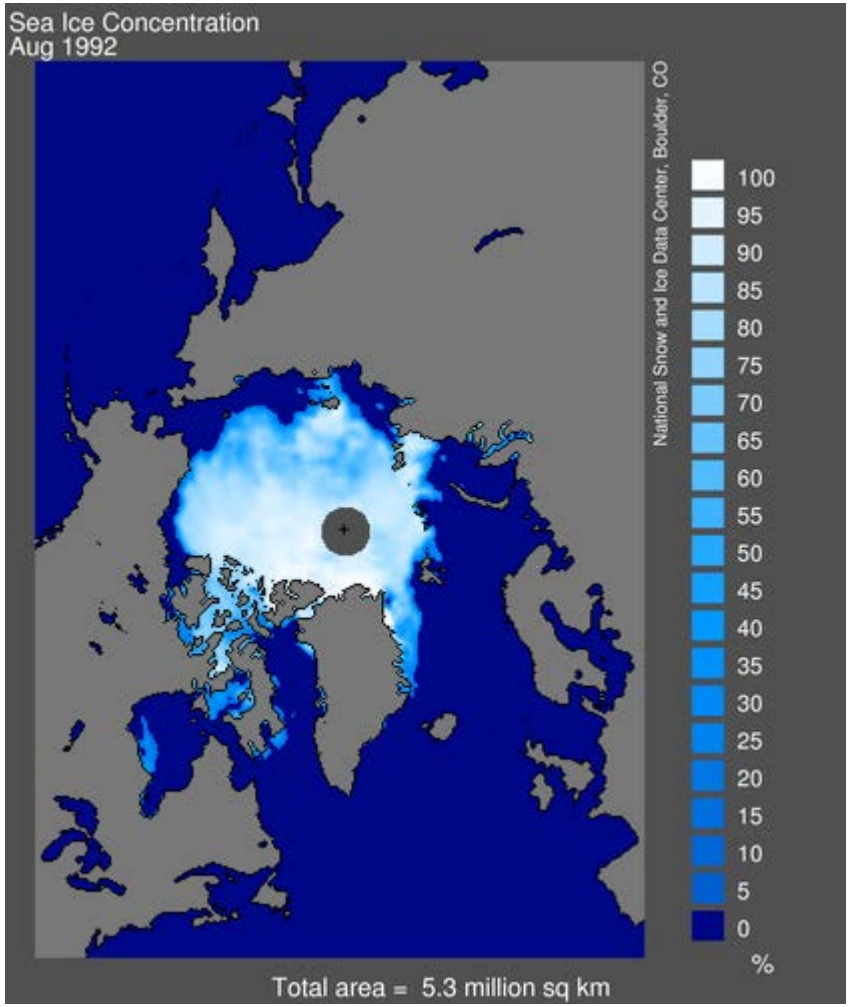
Land surface

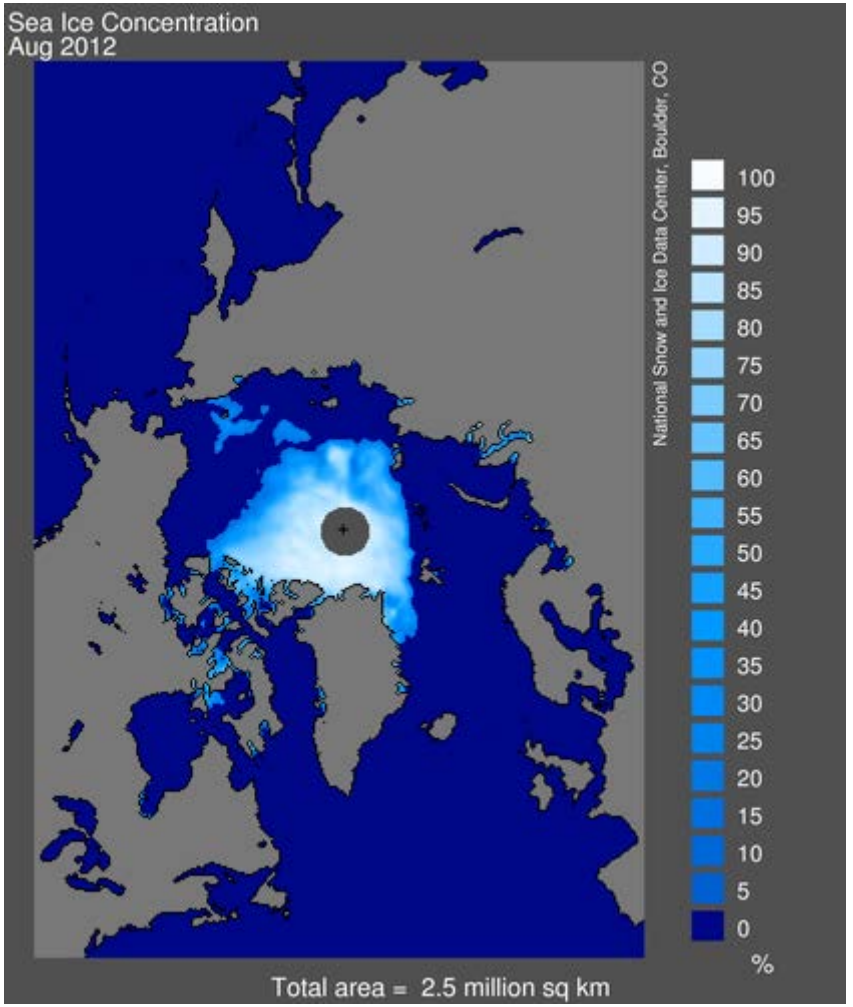




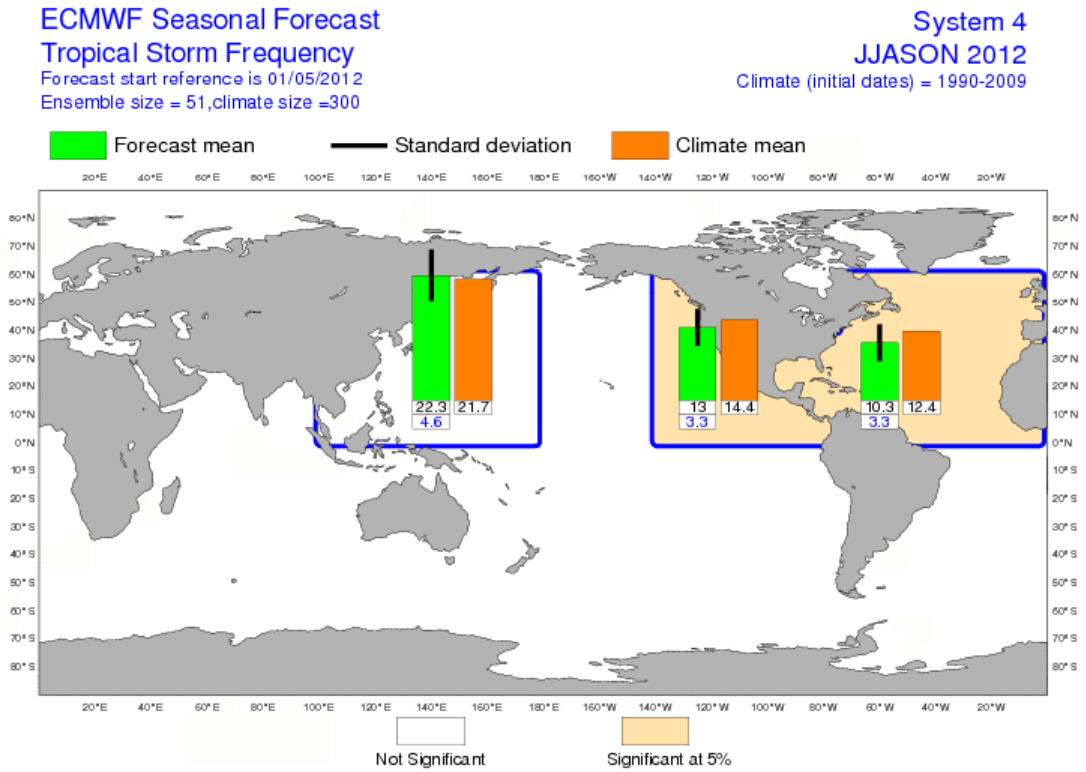
Snow depth limits, 1st April

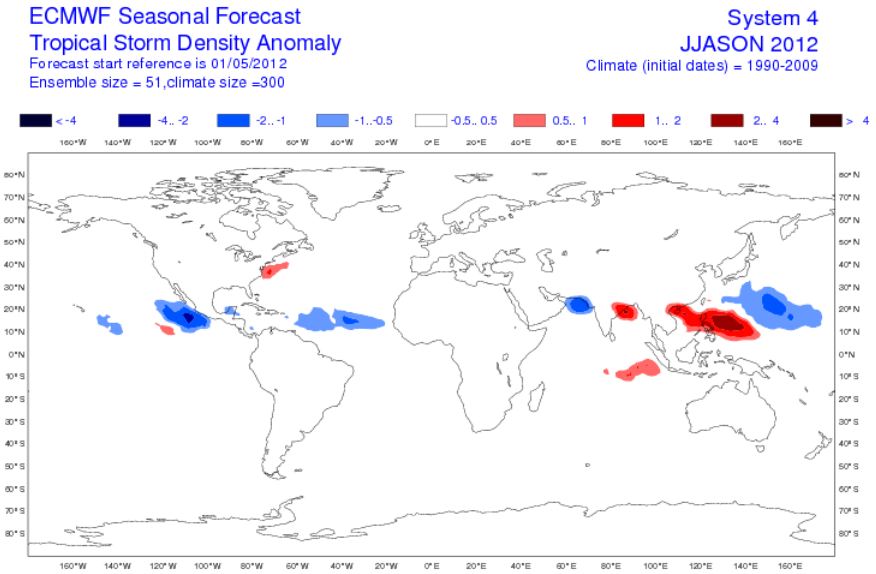
Sea ice

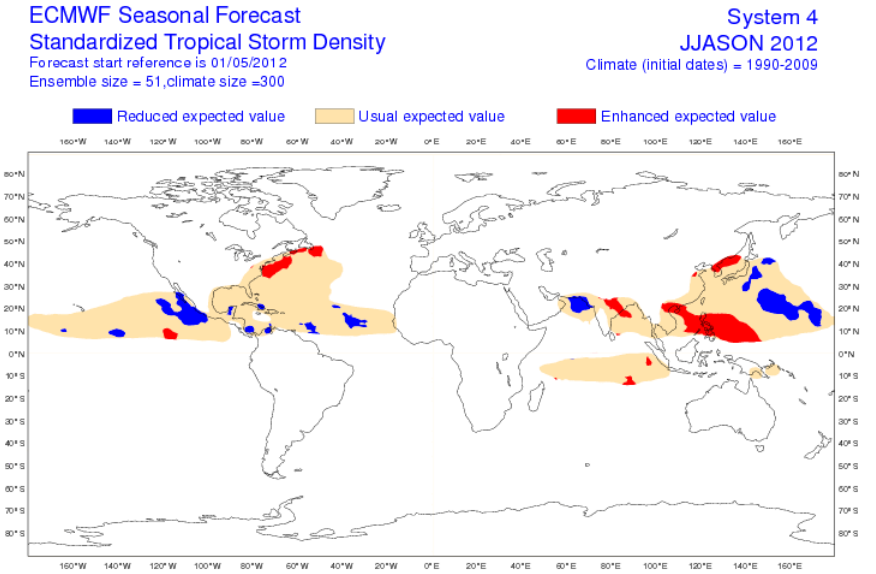

Sea Ice Concentration
Aug 1992
National Snow and Ice Data Center, Boulder, CO
100
95
90
85
80
75
70
65
60
55
50
45
40
35
30
25
20
15
10
5
0
%
Total area = 5.3 million sq km


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



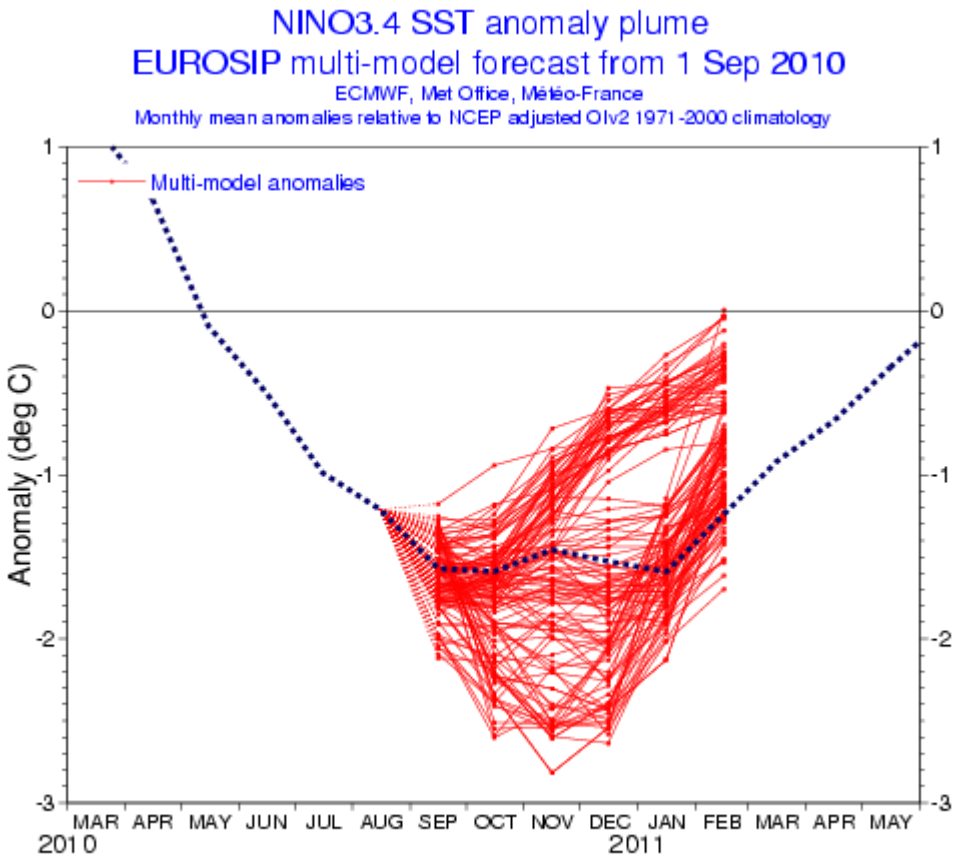




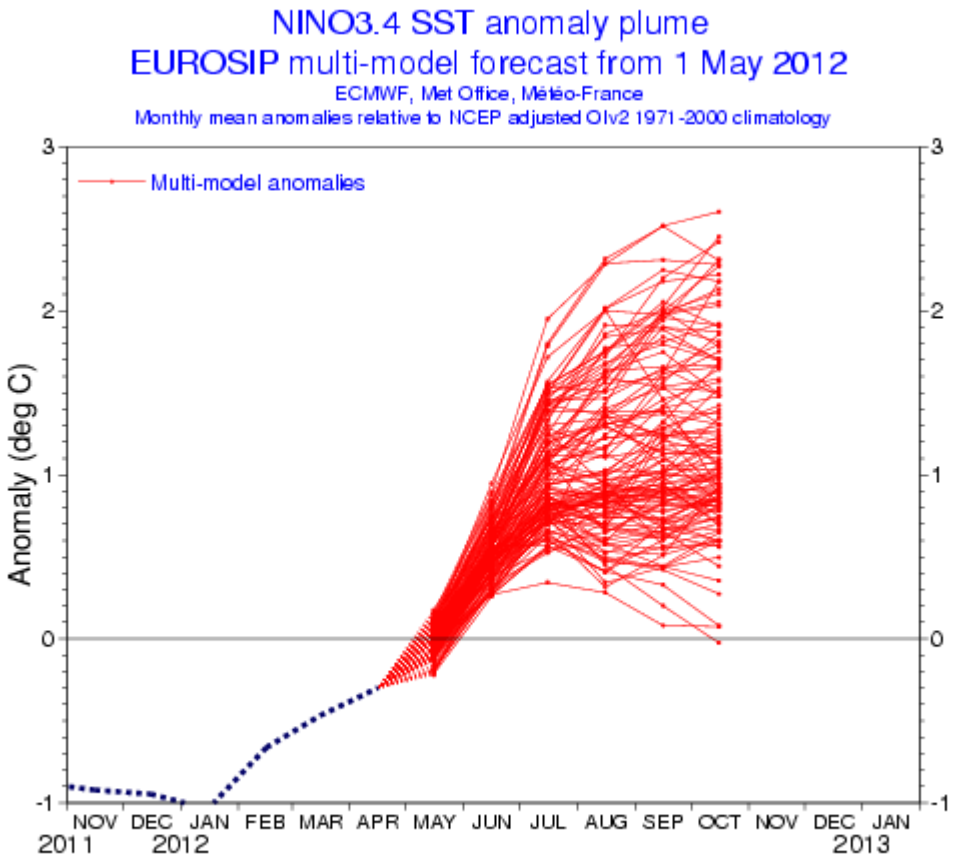
EUROSIP

- **EUROSIP initial design**
  - Co-ordinated forecast strategy
  - Data archive
  - Real-time forecast products
- **Implementation**
  - Initial partners: ECMWF, Met Office, Météo-France
  - Operational from 2005

EUROSIP web products

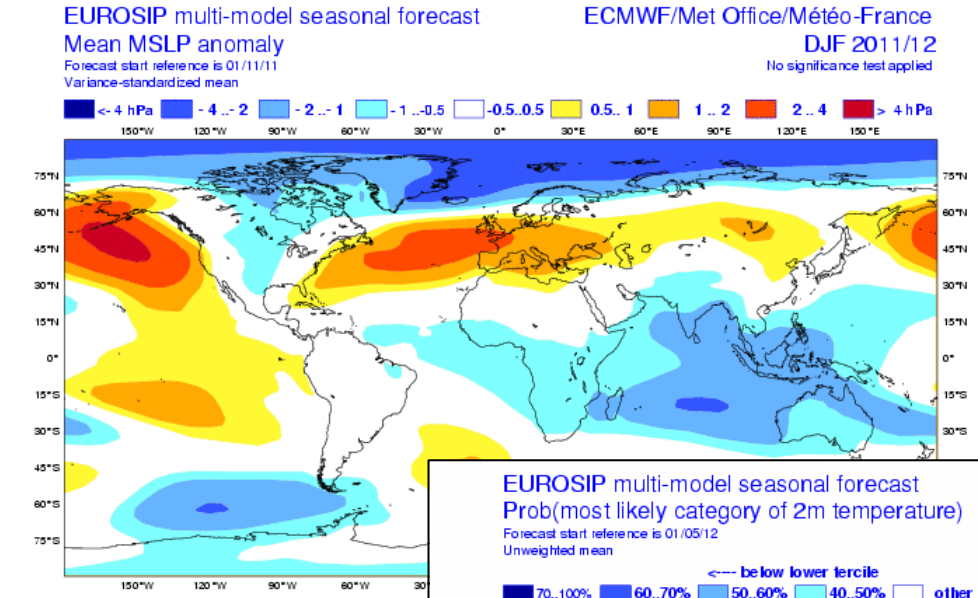


Forecast issue date: 15 Sep 2010

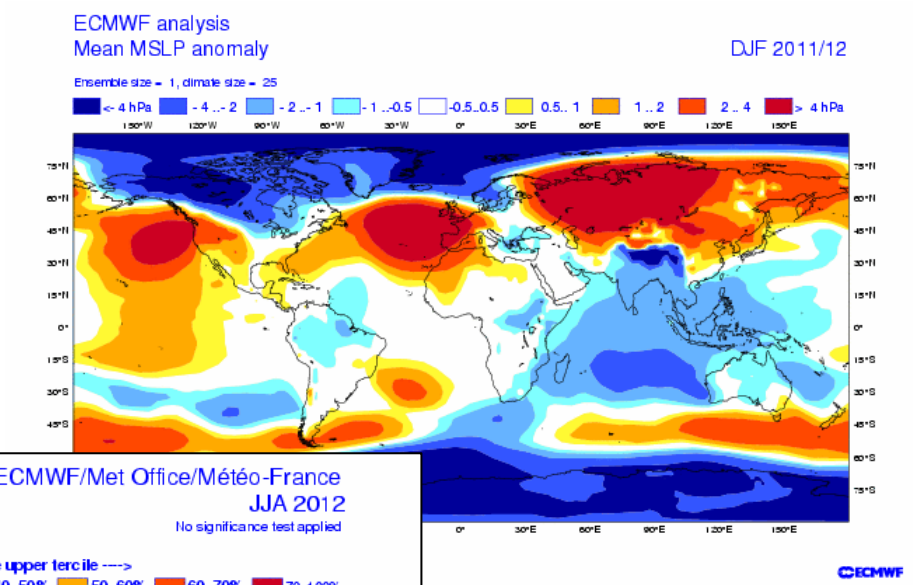


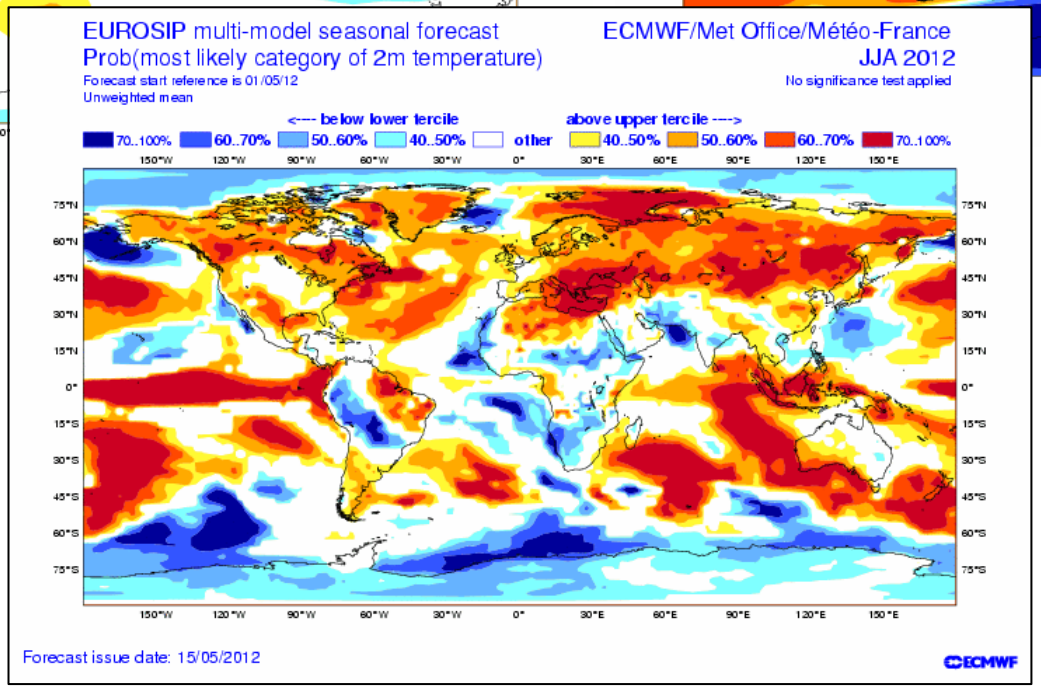
Forecast issue date: 15 May 2012

EUROSIP web products



Forecast issue date: 15/11/2011

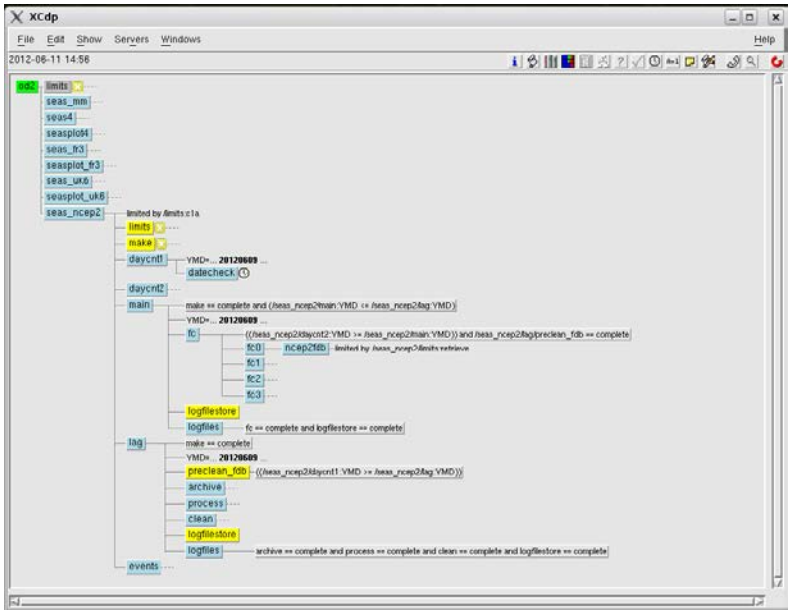




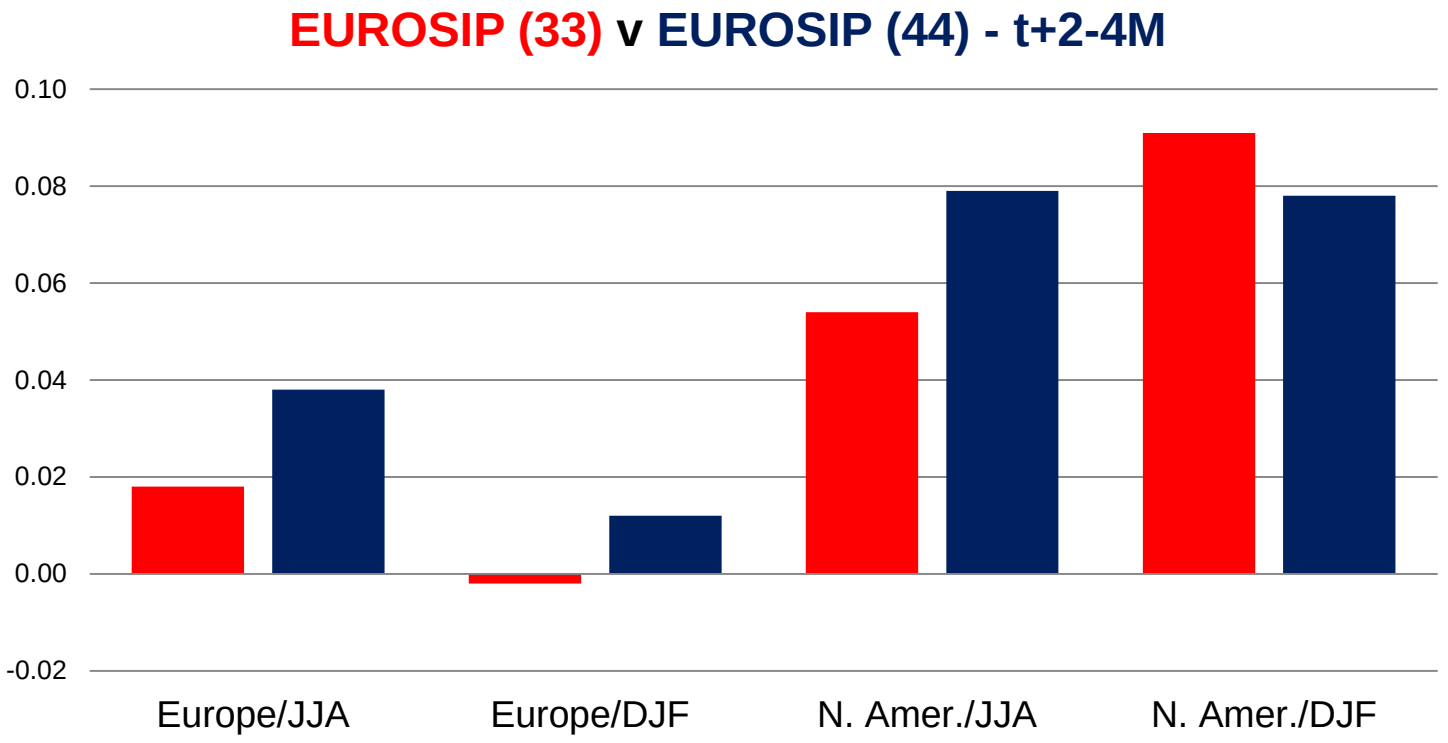
EUROSIP data

- **Individual model data archived in MARS**
  - Daily and monthly means
  - Available to Member States for official duty use
  - Available for research and education
- **Multi-model data products**
  - Created and archived in MARS
  - Available for dissemination, also for commercial customers
- **International support**
  - WMO access to multi-model web products
  - Multi-model data supplied to EUROBRISA project in Brazil

2012: NCEP joins EUROSIP revised processing



Brier Skill Scores (14 years)



ENSO performance

Revised Nino plumes

Variance scaling

- **Robust implementation**
  - Limit to maximum scaling (1.4)
  - Weakened upscaling for very large anomalies
- **Improves *every* individual model**
- **Improves consistency between models**
- **Improves accuracy of multi-model ensemble mean**

Error vs spread

Nino 3.4 plume and pdf

Method for p.d.f. estimation (1)

- **Assume underlying normality**

- **Calculate robust skill-weighted ensemble mean**
  - Do not try a multivariate fit (very small number of data points)
  - Weights estimated ~1/(error variance). Would be optimal for independent errors – i.e., is conservative.
  - Then use 50% uniform weighting, 50% skill dependent

- **Comments:**
  - Rank weighting also tried, but didn't help.
  - QC term tried, using likelihood to downplay impact of outliers, but again didn't help. Outliers are usually wrong, but not always.
  - Models usually agree reasonably well, and tweaks to weights have very little impact anyway.

Method for p.d.f. estimation (2)

- **Re-centre lower-weighted models**
  - To give correct multi-model ensemble mean
  - Done so as to minimize disturbance to multi-model spread
- **Compare past ensemble and error variances**
  - Use above method (cross-validated) to generate past ensembles
  - Unbiased estimates of multi-model ensemble variance and observed error variance
  - Scale forecast ensemble variance
  - 50% of variance is from the scaled climatological value, 50% from the scaled forecast value
- **Comments:**
  - For multi-model, use of predicted spread gives better results
  - For single model, seems not to be so.

Method for p.d.f. estimation (3)

- **Estimate t distribution**
  - Variance estimates are based on small samples, ~15 points
  - Need to use ‘t’ distribution to estimate resulting p.d.f.
  - Finite d.o.f. due to both number of years and ensemble size

- **Plot p.d.f.**
  - Specified percentiles, or plume with 2%ile intervals
  - Or plot forecast values with calibrated mean and variance

- **Comments:**
  - Can apply to single model or multi-model
  - Small ensemble size -> large width of p.d.f.

P.d.f. interpretation

- **P.d.f. based on past errors**
  - The risk of a real-time forecast having a new category of error is not accounted for. E.g. Tambora volcanic eruption.
  - We plot 2% and 98%ile. Would not go beyond this in tails.
  - Risk of change in bias in real-time forecast relative to re-forecast.
- **Bayesian p.d.f.**
  - Explicitly models uncertainty coming from errors in forecasting system
  - Two different systems will calculate different pdf's – both are correct
- **Validation**
  - Rank histograms show pdf's are remarkably accurate (cross-validated)
  - Verifying different periods shows relative bias of different periods can distort pdf – sampling issue in our validation data.

Forecast from 1st August

Operational considerations

- **Quality control**
  - Experience shows that a wide variety of problems and errors can occur
  - Balance between automatic and manual QC
- **Timetable**
  - Multi-model products issued at 12Z on the 15th , without fail
  - Contributor data due earlier, but can be late.
  - Safety margin allows some lateness, plus detection of problems and opportunity to re-send data
  - Option to exclude a model in real time, if missing or corrupted
  - Need to allow for weekends/holidays/system downtime at ECMWF.

Future development of EUROSIP

- **Better individual models**
  - **Météo-France** have a new system running, due to become operational imminently.
  - **Met Office** will introduce a new high-resolution system this November
  - Longer term, models will continue to be refined and extended
- **More models**
  - DWD are working to develop an operational seasonal forecast system to contribute to EUROSIP
  - Other Centres are interested in joining
- **More advanced products**

The latest forecast …..