Seasonal Forecasting at ECMWF

Tim Stockdale
European Centre for Medium-range Weather Forecasts
t.stockdale@ecmwf.int

Contents

Sources of seasonal predictability

- KNOWN TO BE IMPORTANT:
  - El Nino variability - biggest single signal
  - Other tropical ocean SST - important, but multifarious
  - Climate change - impact is substantial, especially in temperature forecasts, and must be accounted for
  - Local land surface conditions - e.g. soil moisture in spring: dry soil warms up more quickly and is more prone to drought

- OTHER FACTORS:
  - Volcanic eruptions - definitely important for large events, gives global cooling plus sometimes a winter warming in parts of the northern hemisphere
  - Mid-latitude ocean temperatures - Complicated story, could be important in some regions
  - Remote soil moisture/ snow cover - Unclear how large the effects might be
  - Sea ice anomalies - definitely local effects, possibly weaker remote impacts
  - Dynamic memory of atmosphere - most likely for first 1-2 months
  - Stratospheric influences - polar vortex, solar cycle, QBO, ozone, …

- Unknown or Unexpected - ???

Methods of seasonal forecasting

- Empirical forecasting
  - Use past observational record and statistical methods
  - Works with reality instead of error-prone numerical models
  - Limited number of past cases means that it works best when observed variability is dominated by a single source of predictability
  - A non-stationary climate is problematic
- Two-tier forecast systems
  - First predict SST anomalies (ENSO or global; dynamical or statistical)
  - Use ensemble of atmosphere GCMs to predict global response
  - Some people still use regression of a predicted El Nino index on a local variable of interest
- Single-tier GCM forecasts
  - Include comprehensive range of sources of predictability
  - Predict joint evolution of SST and atmosphere flow
  - Includes indeterminacy of future SST, important for prob. Forecasts
  - Model errors are an issue!

2. How to make a numerical seasonal forecast

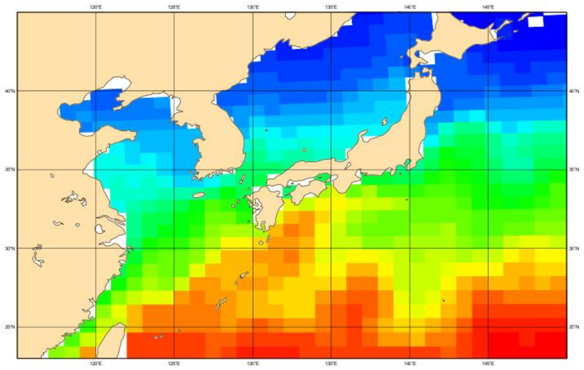
- Step1: Build a comprehensive model
- Step 2: Create initial conditions
- Step 3: Run an ensemble forecast
- Step 4: Calibrate the output

Step 1: Build a comprehensive model

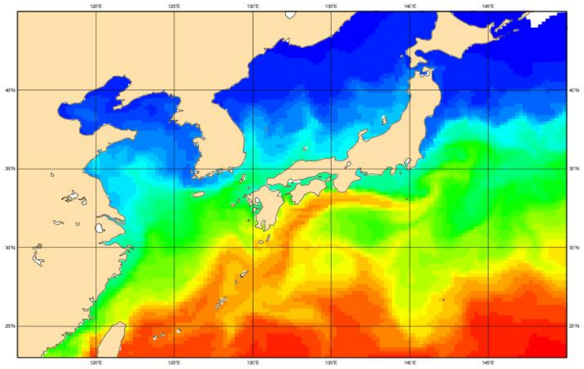
- Aim to have the best model possible, given the amount of computing resource available.
- Include relevant sources of predictability
- ECMWF forecasting model SEAS5, introduced in Nov 2017, includes:
  - IFS (atmosphere and land surface)
    - $T_{CO}319L91$ Cy43r1, 36km grid for physics, full stratosphere
    - All of the physical and dynamic processes of a world-class NWP model
    - Land surface model, multiple soil layers, different soil types, different vegetation types, snow, glaciers
    - Lake model, variable depths, variable mixed layer, surface and bottom temperatures, lake ice
    - Time varying tropospheric sulphate aerosol and stratospheric aerosol from volcanoes
  - Wave model
    - Ocean surface waves modify the interaction between ocean and atmosphere. Runs at 0.5 deg resolution.
  - NEMO (ocean)
    - Global ocean model, 0.25 deg resolution (eddy permitting), 75 vertical levels
  - LIM (sea-ice)
    - Single category ice, solved on same grid as ocean model

SEAS5 – ocean component

- Ocean model resolution upgraded from previous 1x1 deg to 0.25x0.25 deg
- Ocean vertical resolution improved from previous 42 levels to 75 levels
- High ocean resolution is needed to represent ocean eddies, and to better resolve the boundary currents that are important in the ocean, such as the Kuroshio in the Pacific (shown here) and the Gulf Stream in the Atlantic.



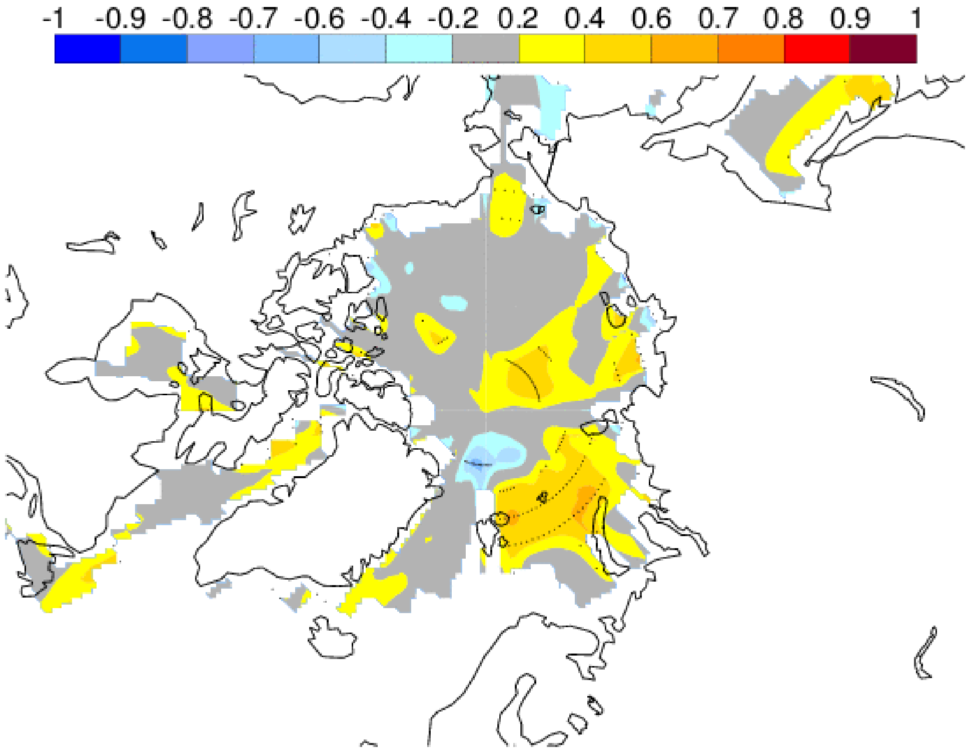
NEMO ORCA1 Z42



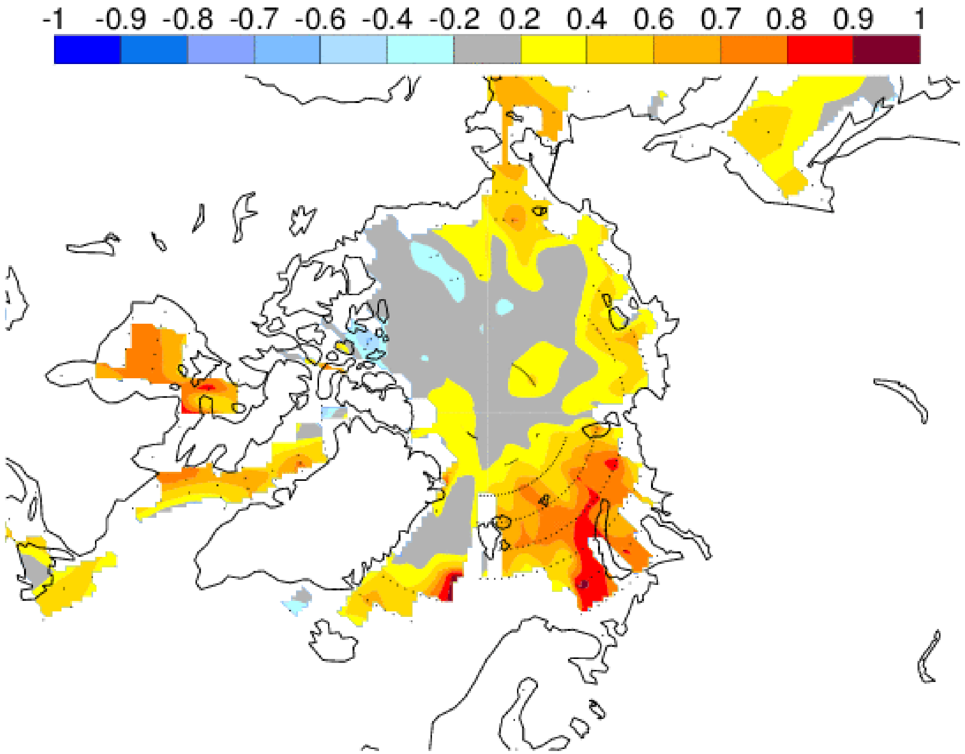
NEMO ORCA025 Z75

SEAS5 - Sea ice model

SEAS4



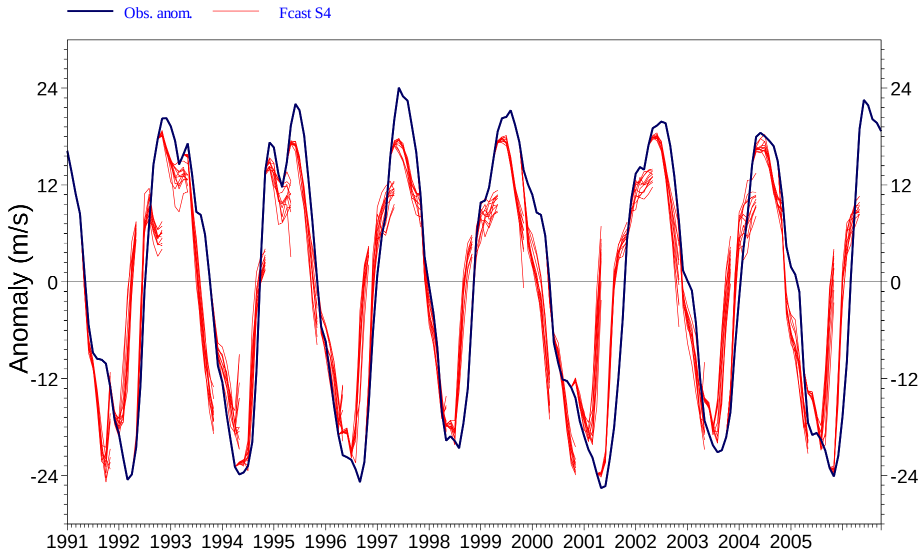
SEAS5



Sea ice cover predictability is improved when we include the interactive sea ice model, illustrated here with correlation scores for predictions of DJF sea ice cover.

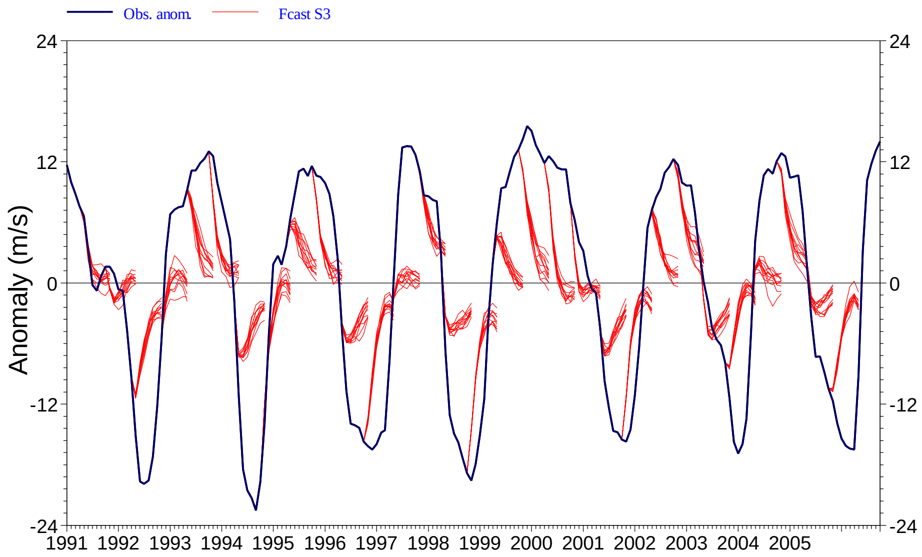
Stratosphere - the QBO

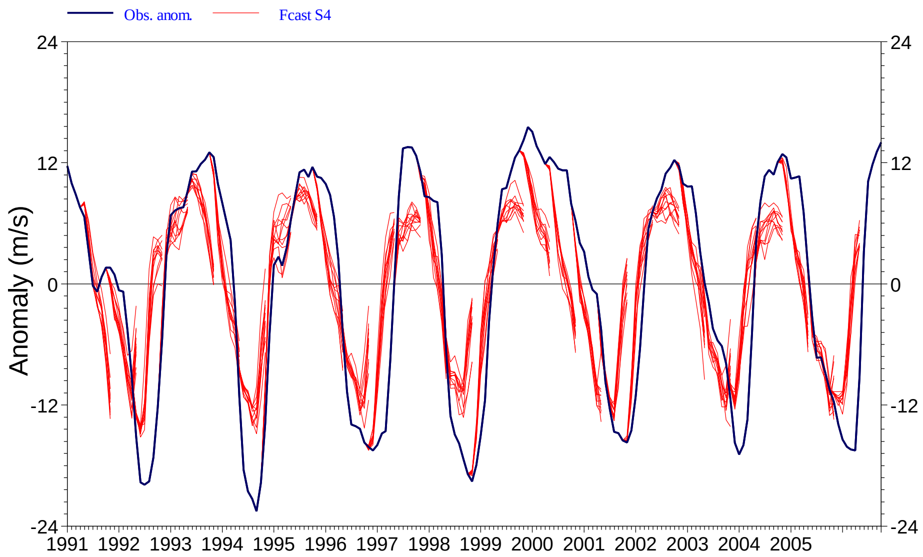
System 4



30hPa

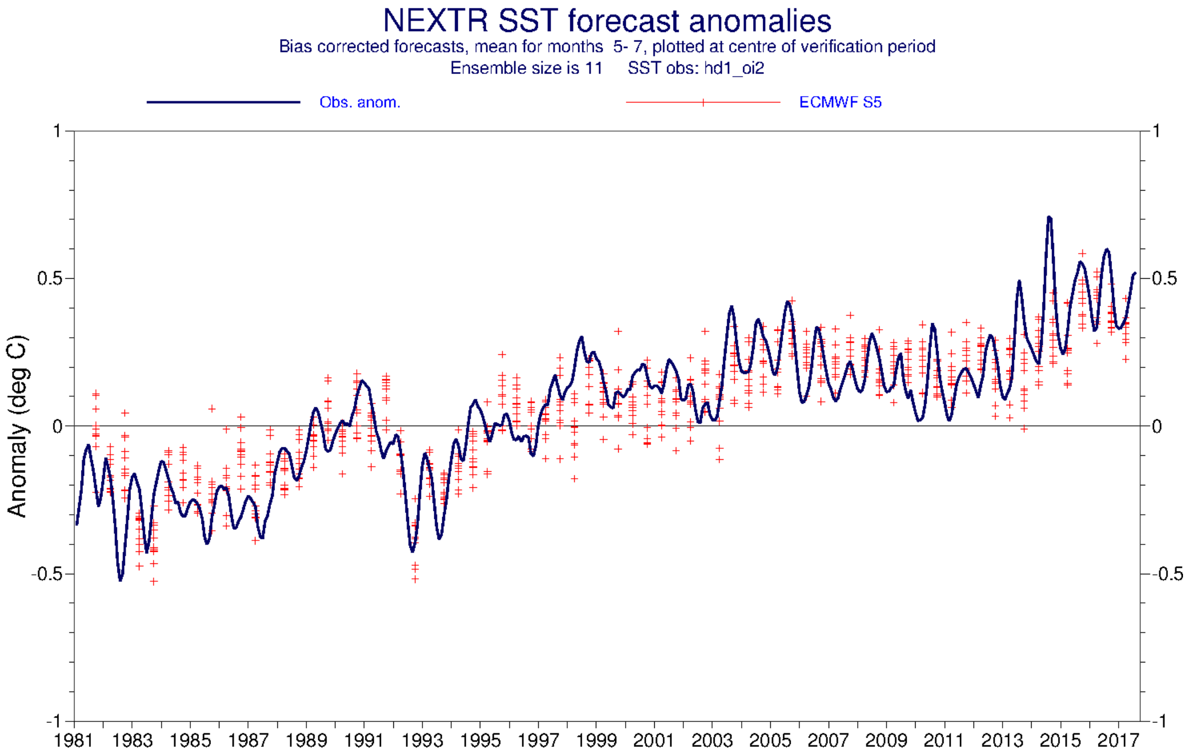
System 3





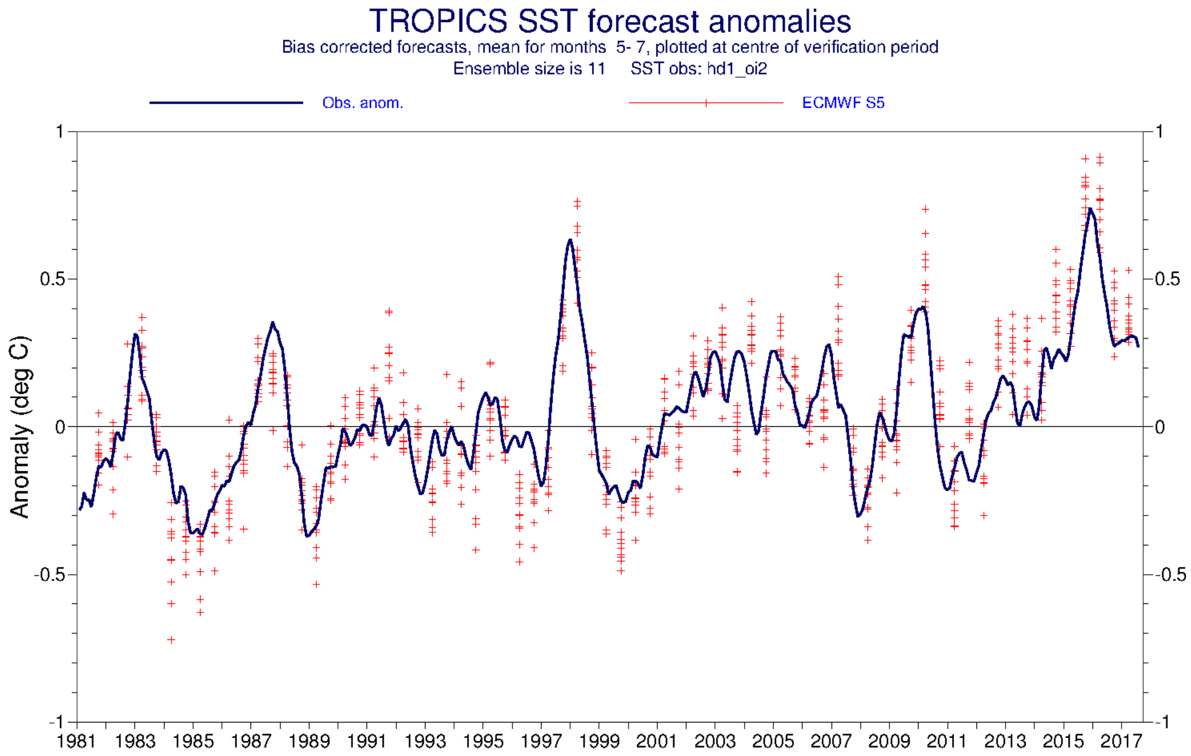
50hPa

Trends

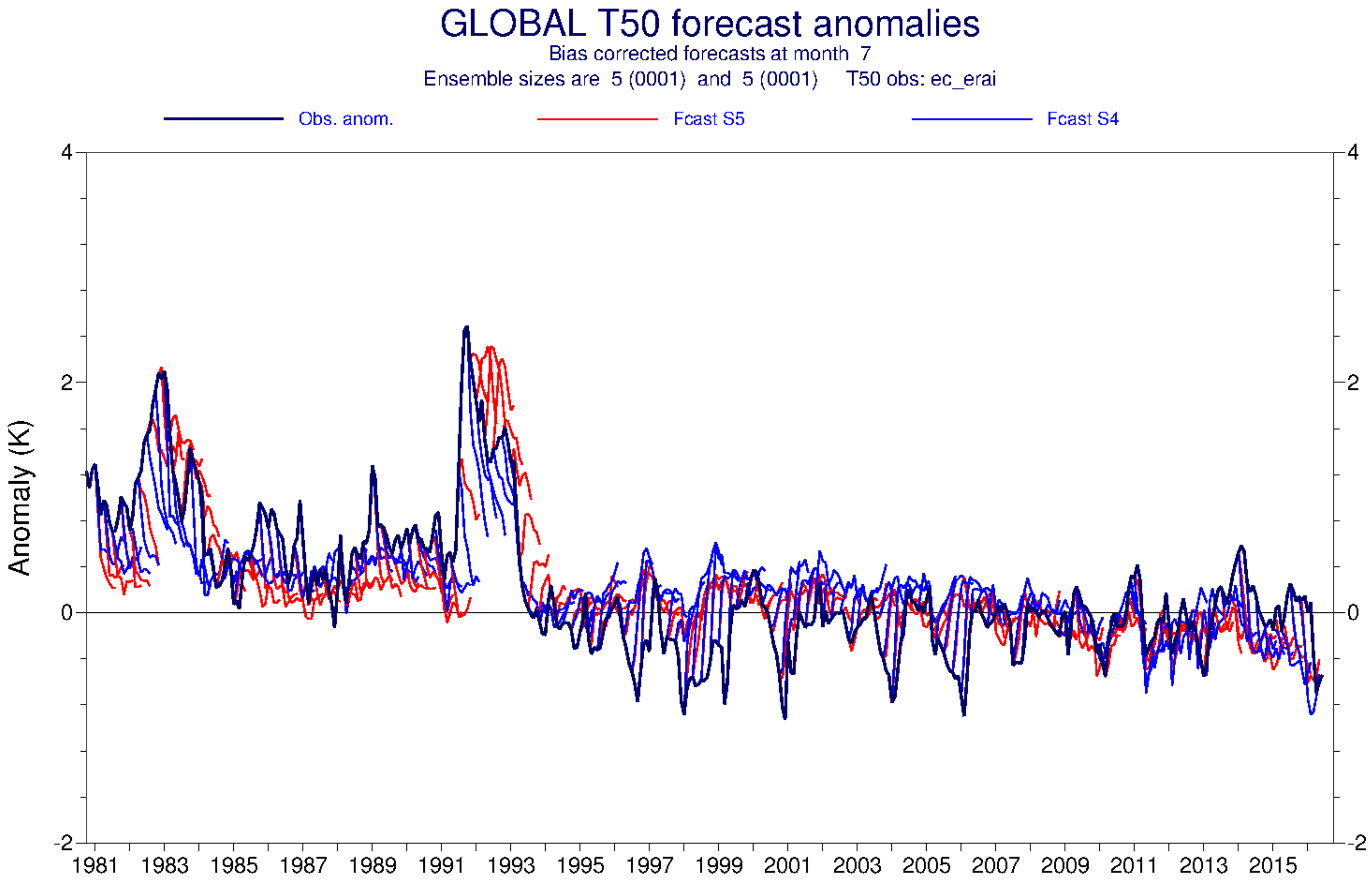


Capturing trends is important. Time-varying CO_2 and other factors are important in this.

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



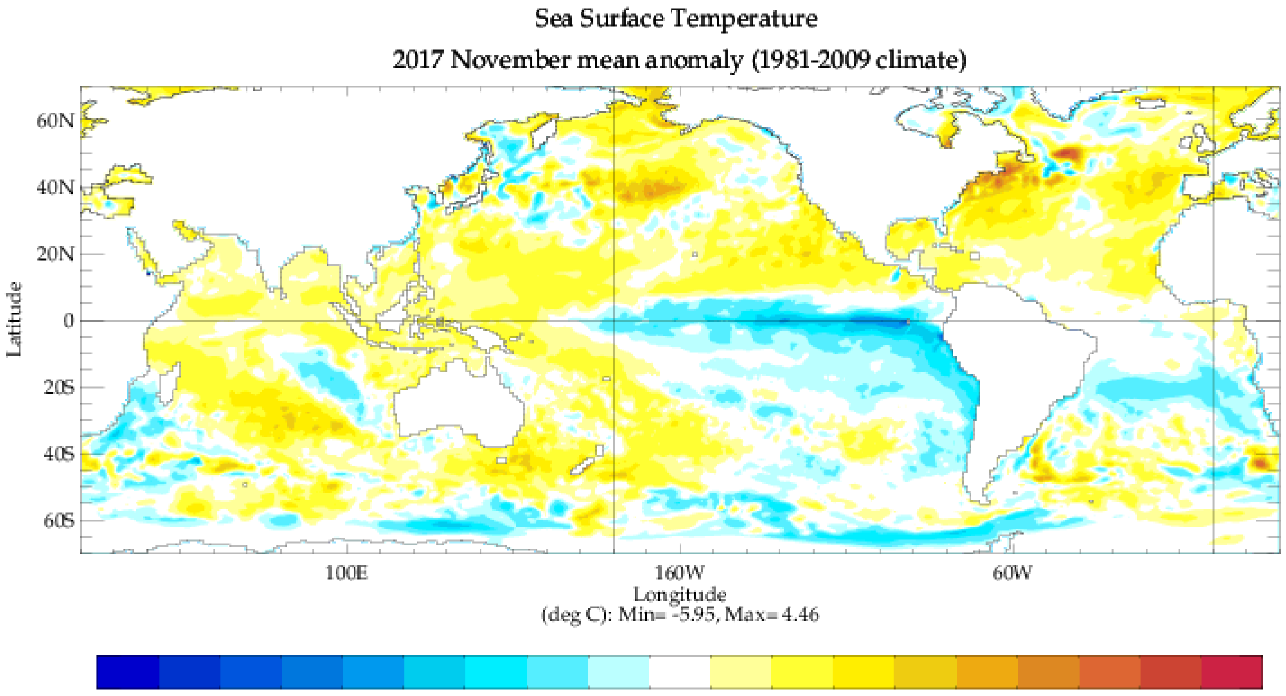
Stratosphere – volcanic aerosols

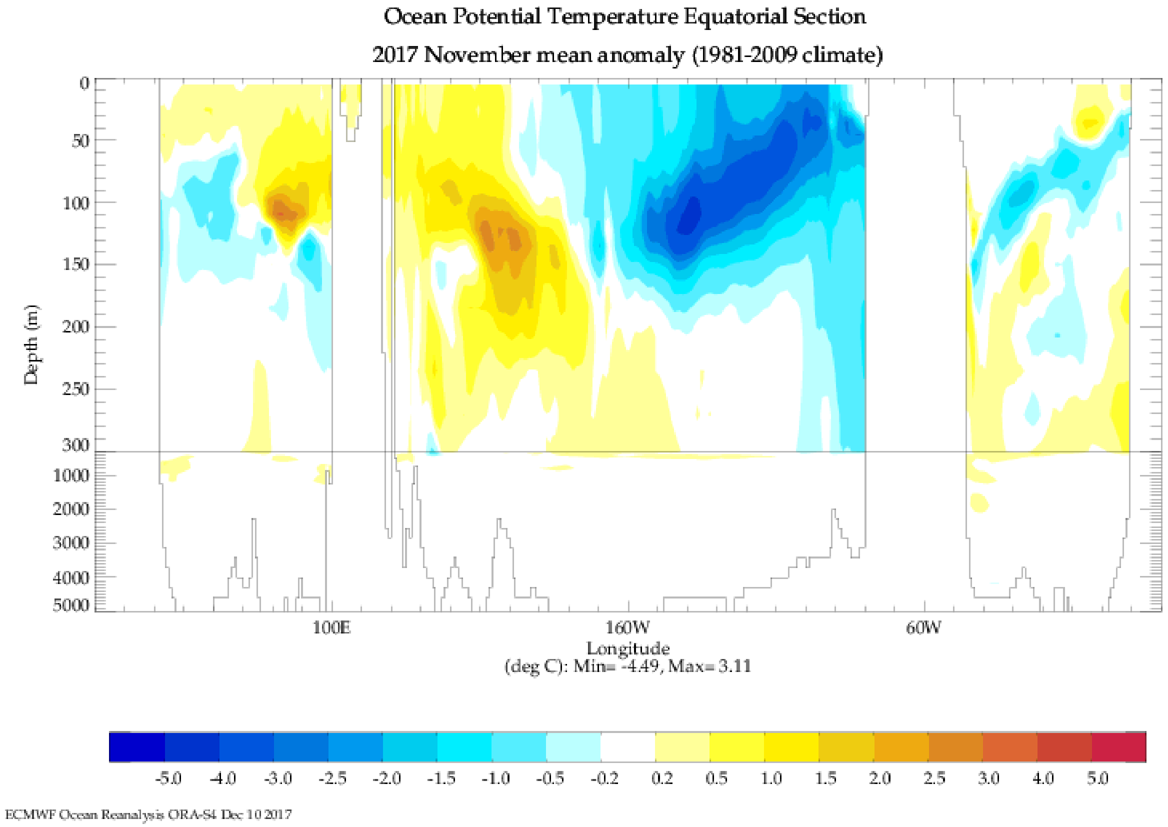


Step 2: Create initial conditions

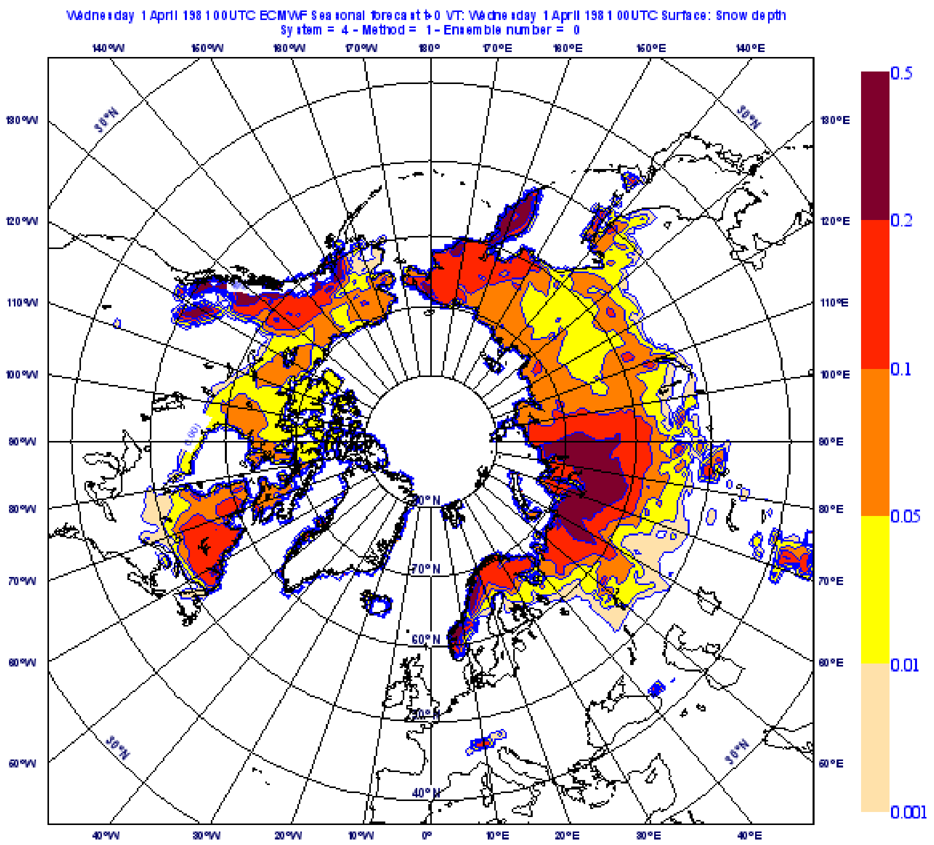
- Most of the forecast skill comes from initial conditions
  - Small amount from changes in forcing terms such as CO_2, volcanoes, solar variability
- Slowly-changing parts of the earth system are typically most important
  - Often these are hard to observe (e.g. sub-surface ocean or below the ground), so creating right initial conditions not easy
- Initial conditions use data assimilation: blending the (often very limited) observational data with model calculations to get the best estimate of the initial state
- A good set of initial conditions will include estimates of the uncertainty, so that this can be accounted for in the forecast.
  - We do this better for some components (e.g. the atmosphere) than we do for others (e.g. land surface).
- We need initial conditions in real-time (so that we can run today's forecast), but also for many years in the past to allow for calibration (see later).
  - Initial conditions need to be both accurate and consistent over time for calibration purposes. It is not easy to ensure both of these are true, when data is needed in real-time but also stretching several decades into the past. For example, the ocean is much better observed now than in the past.

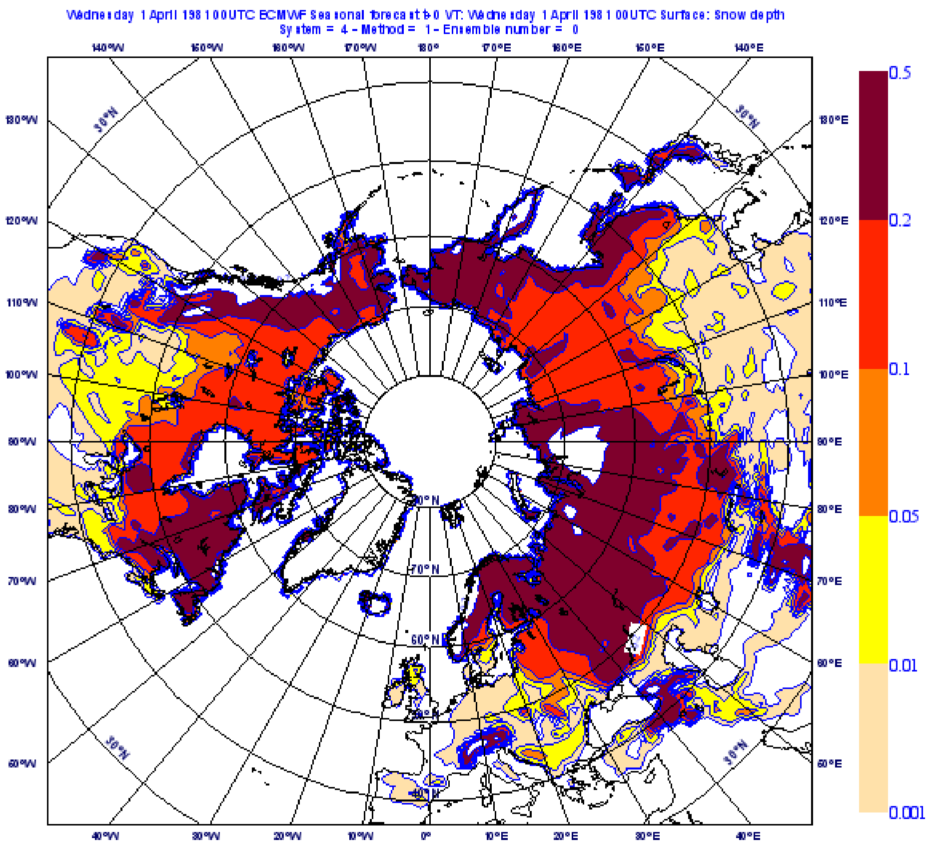
Initial conditions – ocean analysis





Initial conditions - land surface





Snow depth limits, 1st April

Step 3: Run an ensemble forecast

- Need an ensemble to represent possible future evolutions of the atmosphere/ocean/earth system
  - After 10-15 days, details of day-to-day weather become decorrelated – all the ensemble members predict something different, and the exact future is unknowable
  - But we are looking for changes in the pdf of, for example, seasonal means. Ensembles will give us the information we need to evaluate such changes and probabilities.
- How to make the ensemble members different?
  - Explicitly generate an ensemble on the 1st of each month, with perturbations to represent the uncertainty *in the initial conditions*; run forecasts for 7 months
    - SST perturbations (symmetric pairs, based on sampling of past differences between different analyses)
    - 5 member ocean analysis, with perturbations to wind and data
  - Stochastic physics to represent *indeterminacy* of large scale (due to missing details of small scales) and also account for uncertainty due to our model being imperfect
- Now, simply run the ensemble of forecasts
  - This is where we need a very big computer
  - Worry about model biases later ….

ECMWF SEAS5 configuration

- Real time forecasts:
  - **51 member ensemble forecast to 7 months**
  - SST and atmosphere initial perturbations (SV, EDA) added to each member

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

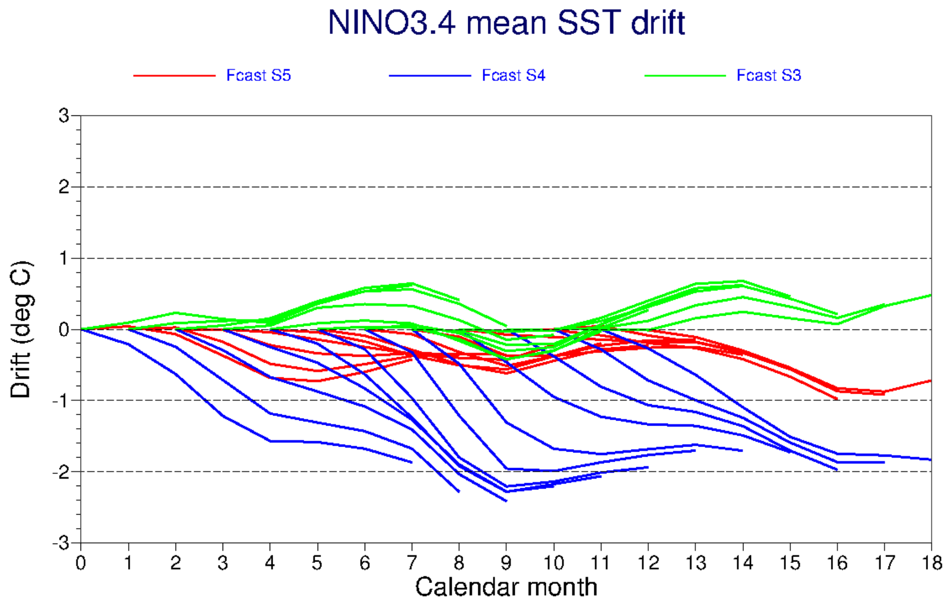
- Re-forecasts from 1981-2016 (36 years)
  - 25 member ensemble every month
  - 15 members extended to 13 months once per quarter

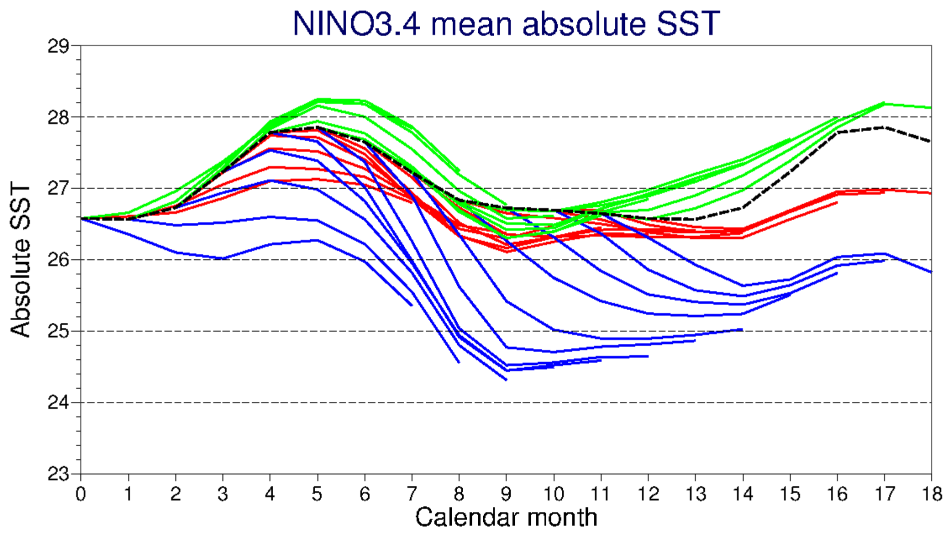
How many re-forecasts?

- Re-forecasts dominate total cost of system
  - SEAS5: 10800 re-forecasts (must be in first year)
    612 real-time integrations (per year)
- Re-forecasts define model climate
  - Need both climate mean and the pdf, latter needs large sample
  - May prefer to use a “recent” period (SEAS5 has 36 years available, but uses only last 24 years for web products)
  - SEAS5 has 600 member climate (25 members * 24 years) for web products, so sampling is basically OK
- Re-forecasts 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 25 member ensemble
  - For regions of high signal/noise, SEAS5 gives adequate skill estimates
  - For regions of low signal/noise (eg <= 0.5), need hundreds of years, 36 years available is not enough

Step 4: Calibrate the output

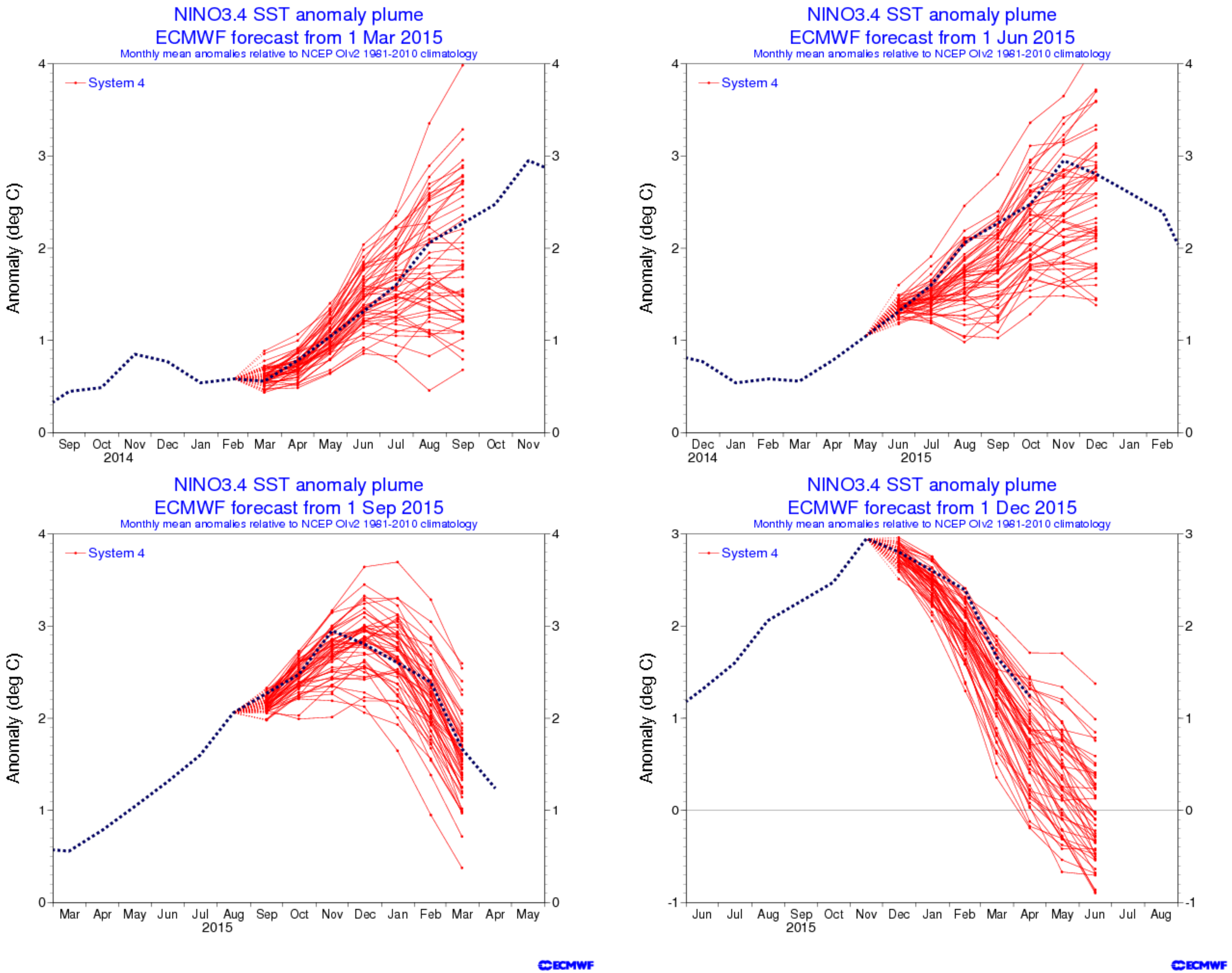
- Model biases are typically comparable in size to the signal we are predicting
  - True for both SST and atmosphere fields
  - This bias **MUST** be accounted for in some way – a fundamental requirement for seasonal forecasting
- Forecast calibration requires a corresponding set of **re-forecasts:**
  - Re-forecasts should use the same model and (where possible) the same method of initialization, to ensure that the biases are consistent
  - There are different ways of using the re-forecasts for calibration, but in general need a large number of re-forecasts. The full set of re-forecasts for a given calendar start date (e.g. 1 May) define the *model climate* for forecasts starting on that date.
- Compare the model forecast to the model climate:
  - SEAS5 forecasts are calibrated using 25-member re-forecasts for each of the years 1993-2016 (24 years), so the model climate has 600 members.
  - Model climate has both a mean and a distribution, allowing us to estimate e.g. tercile boundaries.
  - Model climate is a function of start date and forecast lead time – model biases often get larger the further into the future we calculate.





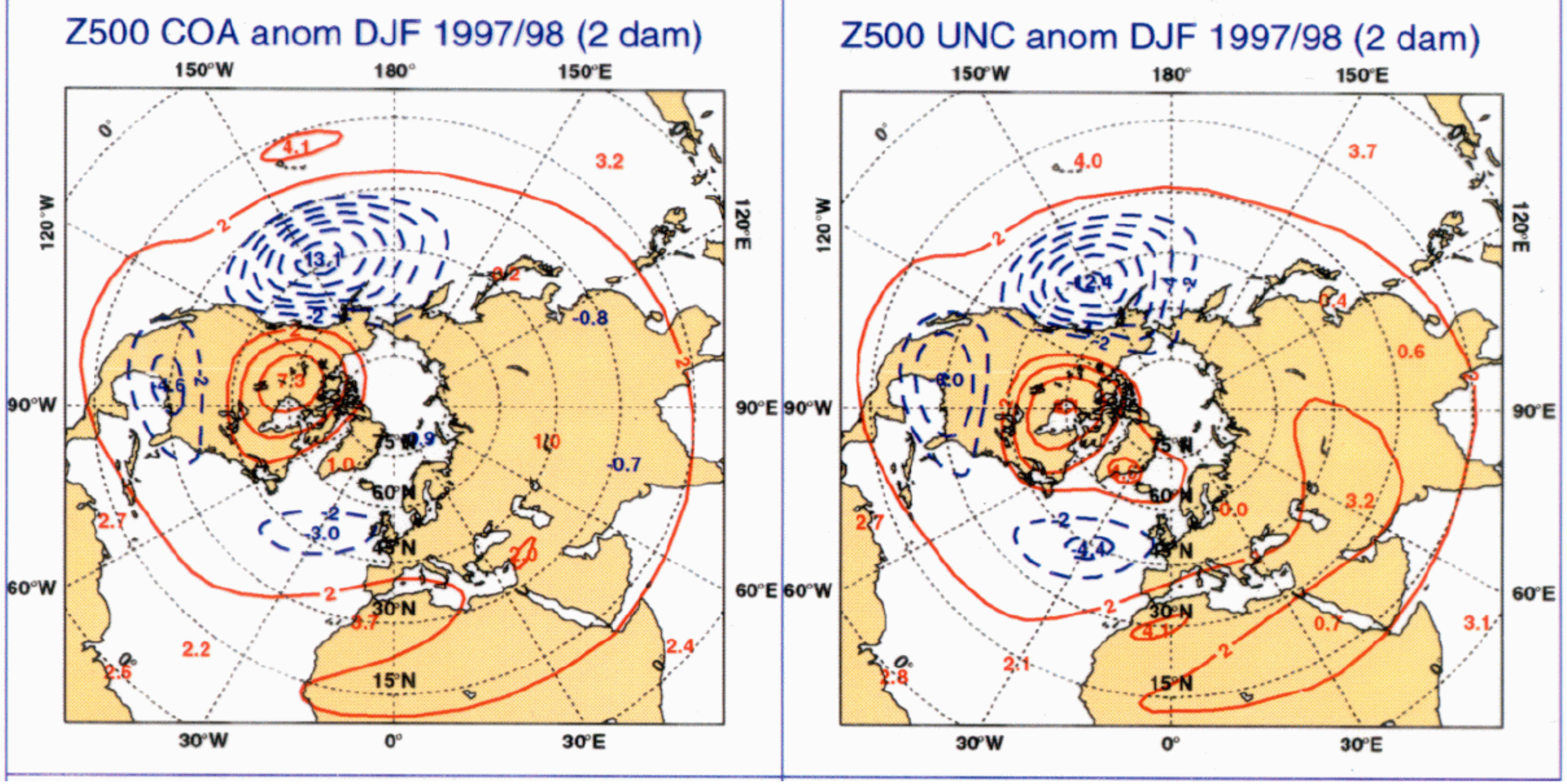
SST bias is a function of lead time and season.

Some systems have less bias, but it is still large enough to require correcting for.

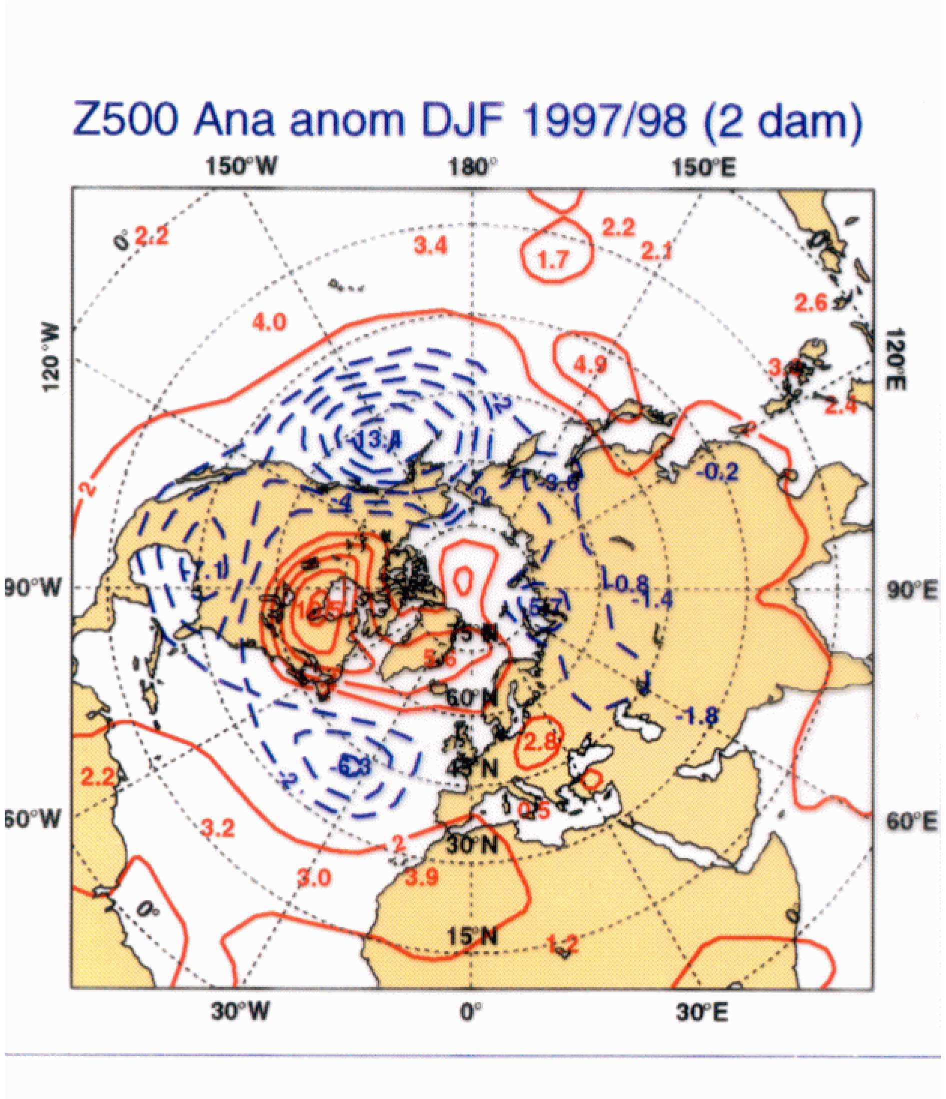


NINO3.4 SST anomaly plume
ECMWF forecast from 1 Mar 2015
Monthly mean anomalies relative to NCEP OIv2 1981-2010 climatology
System 4
Anomaly (deg C)
Sep Oct Nov Dec Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov
2014
NINO3.4 SST anomaly plume
ECMWF forecast from 1 Jun 2015
Monthly mean anomalies relative to NCEP OIv2 1981-2010 climatology
System 4
Anomaly (deg C)
Dec Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec Jan Feb
2014 2015
NINO3.4 SST anomaly plume
ECMWF forecast from 1 Sep 2015
Monthly mean anomalies relative to NCEP OIv2 1981-2010 climatology
System 4
Anomaly (deg C)
Mar Apr May Jun Jul Aug Sep Oct Nov Dec Jan Feb Mar Apr May
2015
NINO3.4 SST anomaly plume
ECMWF forecast from 1 Dec 2015
Monthly mean anomalies relative to NCEP OIv2 1981-2010 climatology
System 4
Anomaly (deg C)
Jun Jul Aug Sep Oct Nov Dec Jan Feb Mar Apr May Jun Jul Aug
2015
ECMWF
ECMWF

Despite SST bias and other errors, anomalies in the coupled system can be remarkably similar to those obtained using observed (unbiased) SSTs …..



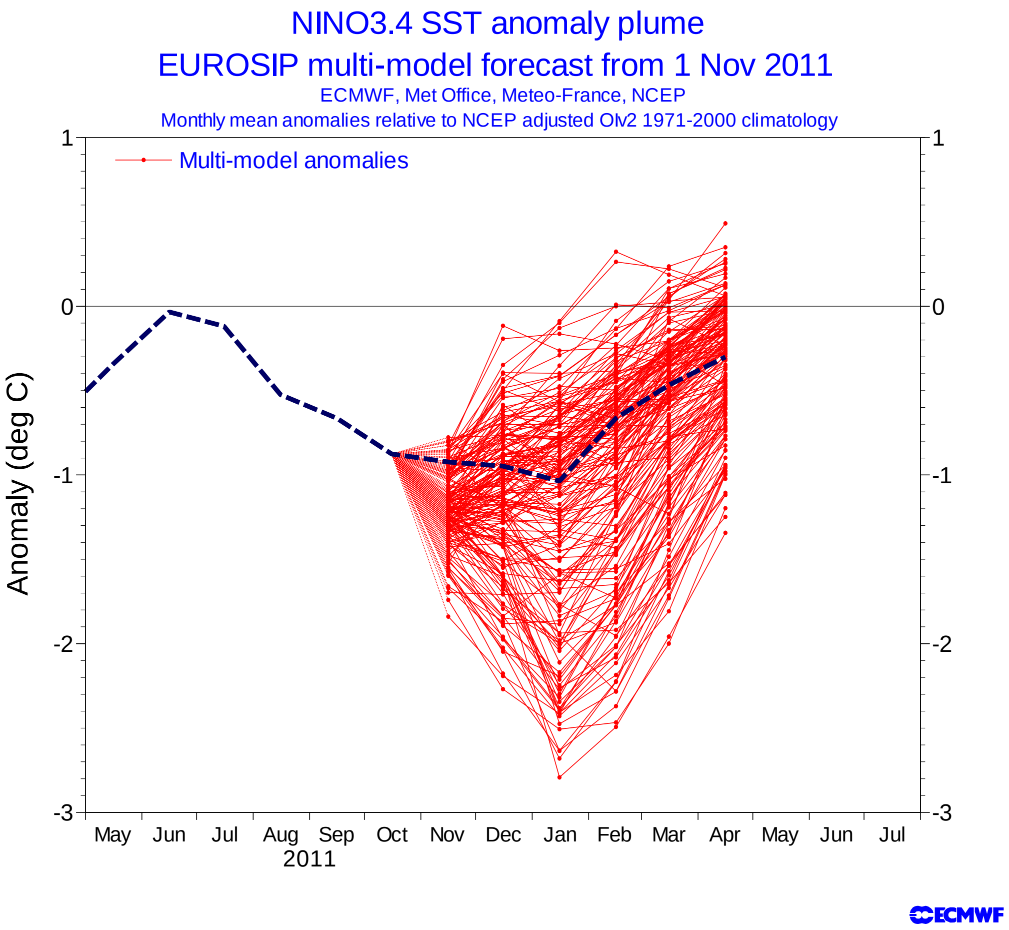
… and can also verify well against observations

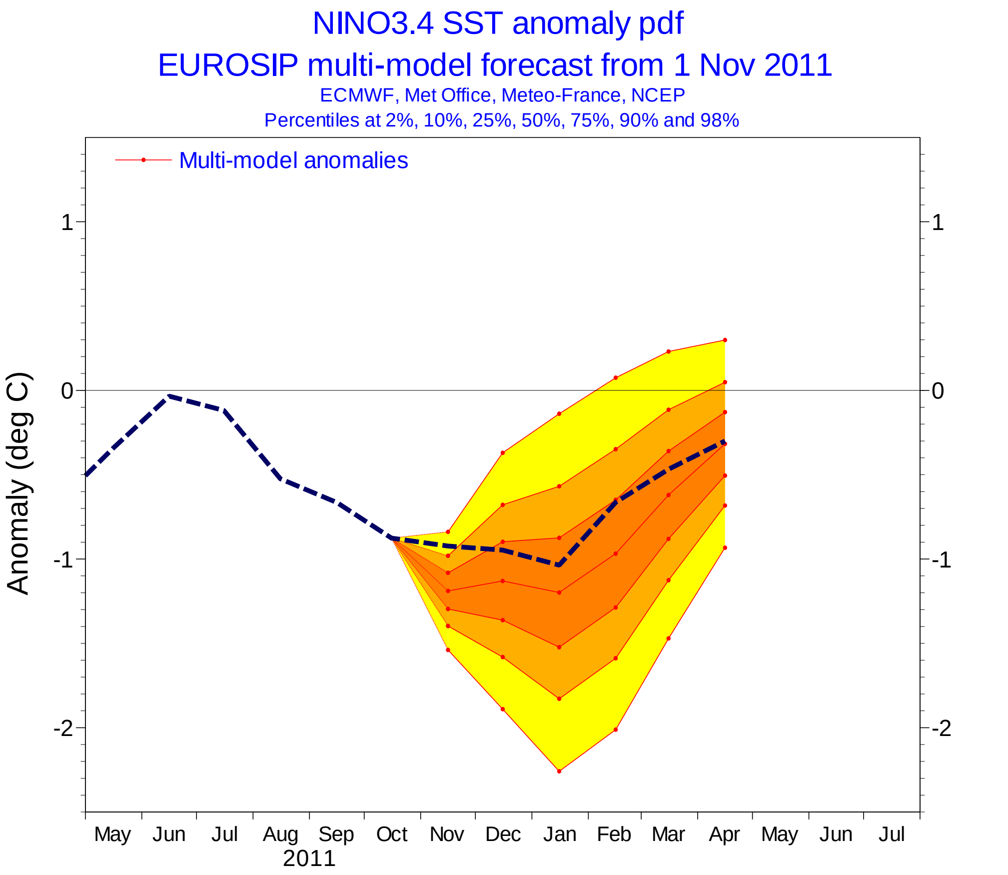


- Implicit assumption of linearity
  - We implicitly assume that a shift in the model forecast relative to the model climate corresponds to the expected shift in a true forecast relative to the true climate, despite differences between model and true climate.
  - Most of the time, this assumption seems to work pretty well. But not always. This is one reason we are always seeking to improve the models and reduce the size of the biases. The smaller the bias, the less chance there is of non-linear effects causing errors in the seasonal forecast.

- More advanced calibration, based on forecast skill
  - The calibration mentioned so far is rather basic, and is designed to remove zeroth or at most first order errors from the forecast
  - Forecast interpretation has to take account of past skill – are there grounds for trusting the model forecast?
  - More advanced calibration methods can be used – we covered some in the previous lecture

Nino 3.4 plume and pdf – calibrated multi-model forecast





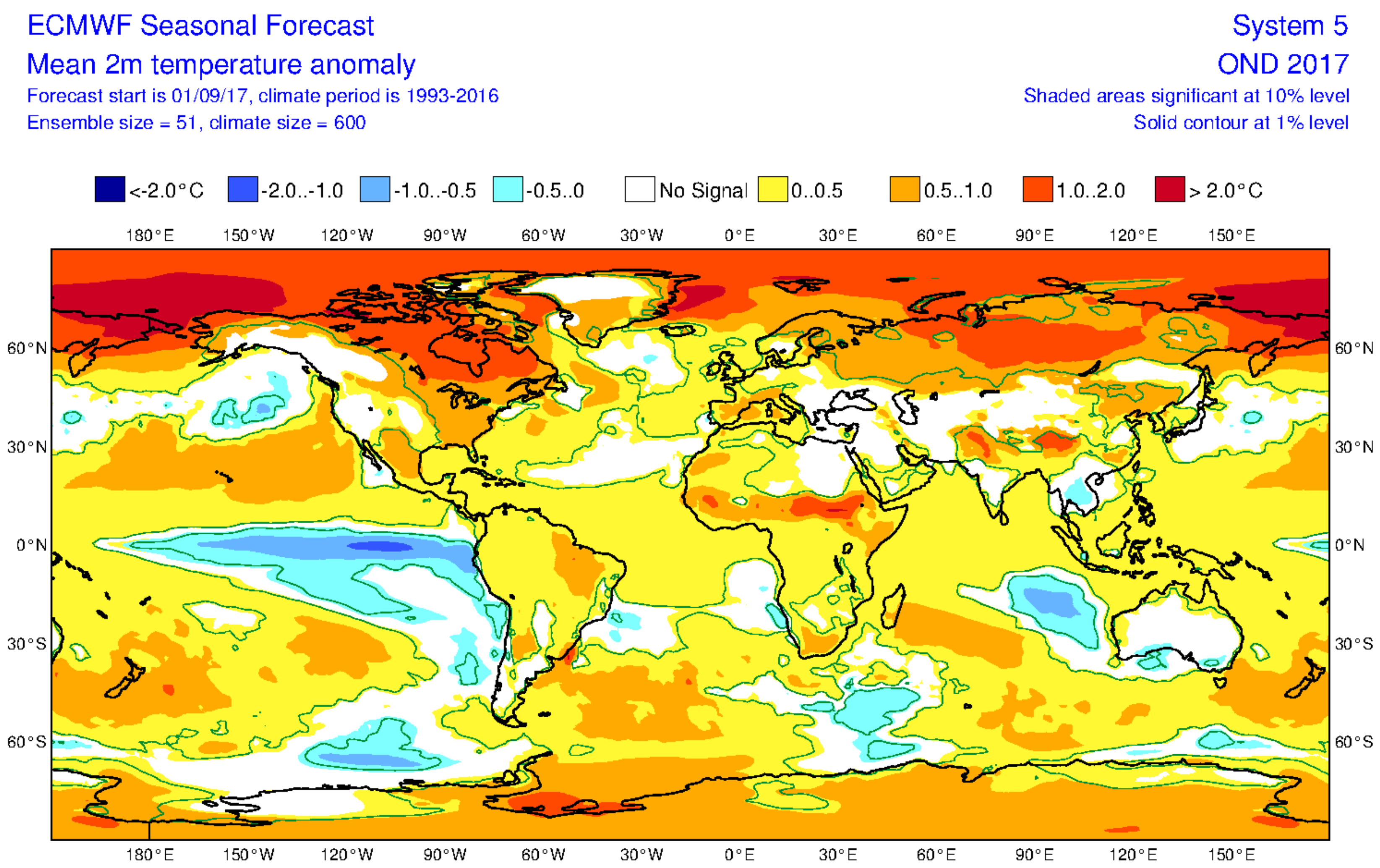
Each model bias and variance corrected, then multi-model ensemble formed, then calibrated to give pdf with correct average width (“t” distribution, variance is half climatological, half varying with model distribution – a variant of NGR)

3. Operational forecast products from SEAS5

- A few examples only – see ECMWF or C3S web pages for full details and assessment of skill

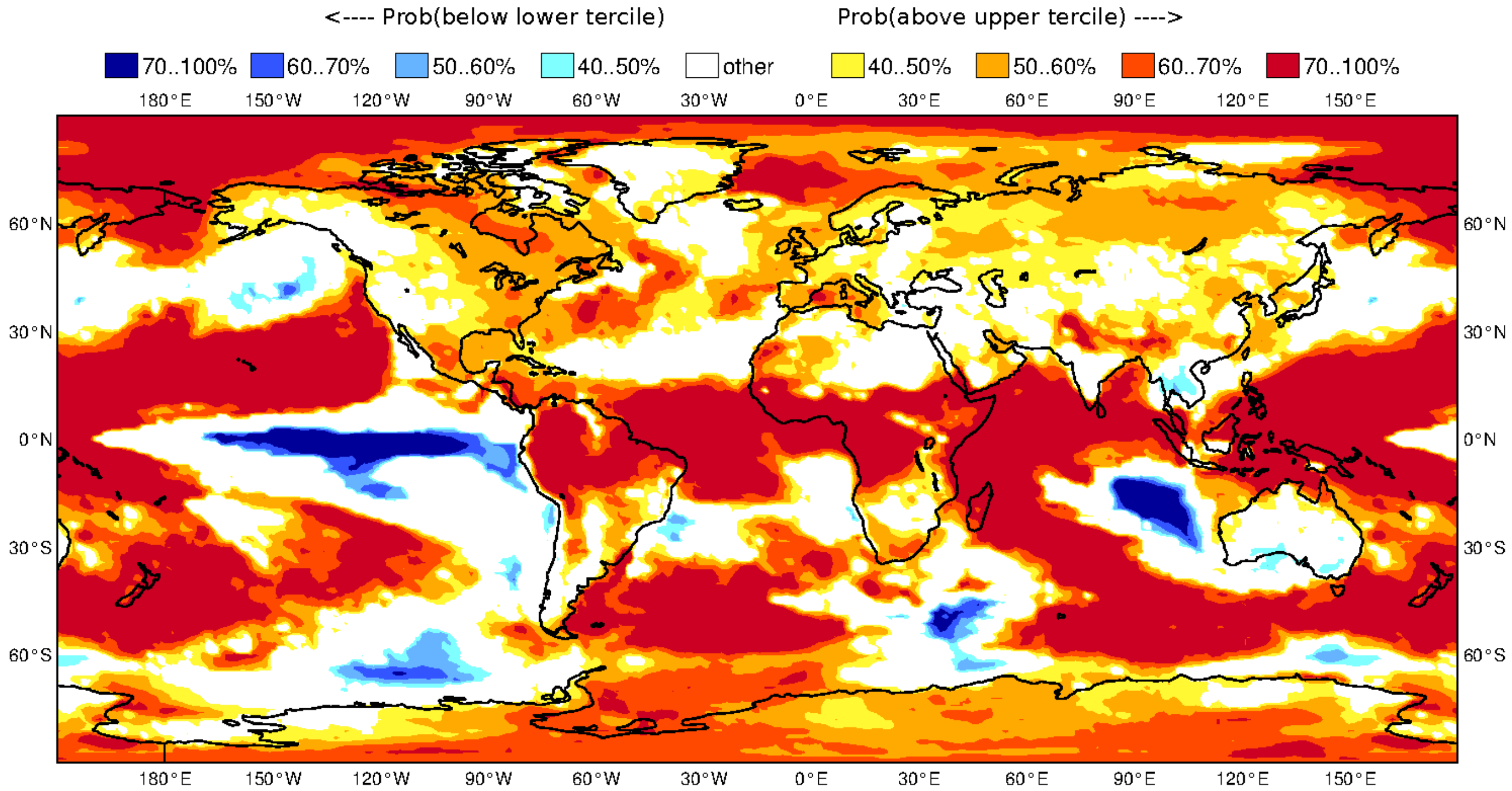
- All graphical products are supplied with corresponding skill estimates, as we illustrate in our examples. It is always important to look at the skill, to reduce the risk of over-interpreting the forecasts!

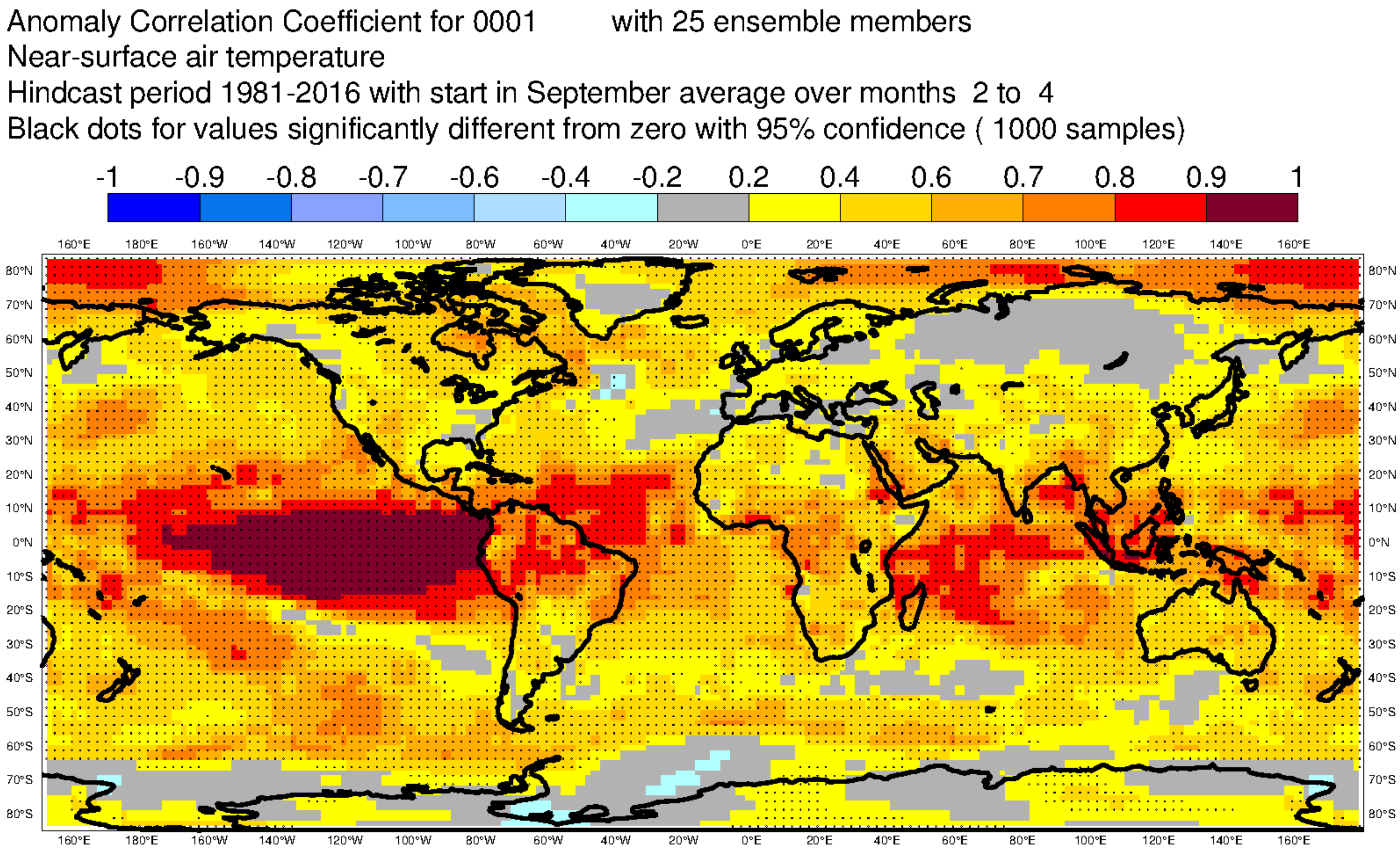
- Note: Significance values on plots
  - Ensembles are large enough to test whether any apparent signals are real shifts in the model pdf, or are due to the limited ensemble size giving a false signal by chance
  - We use the Wilcoxon-Mann-Whitney rank-sum test, which is non-parametric and is both robust and efficient at detecting shifts in the mean
  - The significance levels on the plots are a test as to *whether the model has a signal,* and are NOT related in any way as to whether the model signal should be trusted. The past skill estimates should be looked at to get a sense of the reliability of the model forecasts.


ECMWF Seasonal Forecast
Mean 2m temperature anomaly
Forecast start is 01/09/17, climate period is 1993-2016
Ensemble size = 51, climate size = 600
System 5
OND 2017
Shaded areas significant at 10% level
Solid contour at 1% level
<-2.0°C
-2.0..-1.0
-1.0..-0.5
-0.5..0
No Signal
0..0.5
0.5..1.0
1.0..2.0
> 2.0°C
180°E
150°W
120°W
90°W
60°W
30°W
0°E
30°E
60°E
90°E
120°E
150°E
60°N
30°N
0°N
30°S
60°S

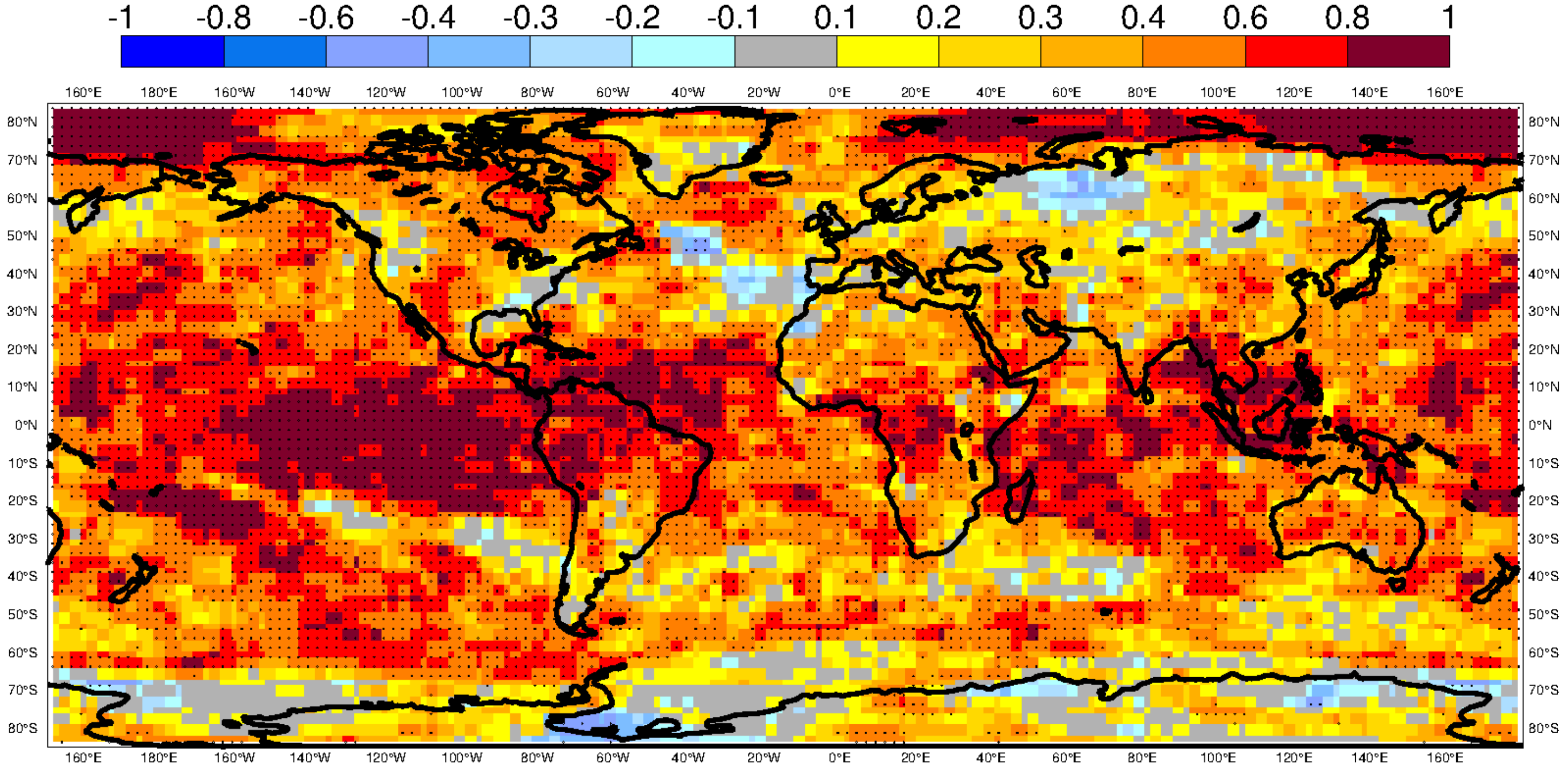
ECMWF Seasonal Forecast
Prob(most likely category of 2m temperature)
Forecast start is 01/09/17, climate period is 1993-2016
Ensemble size = 51, climate size = 600

System 5
OND 2017

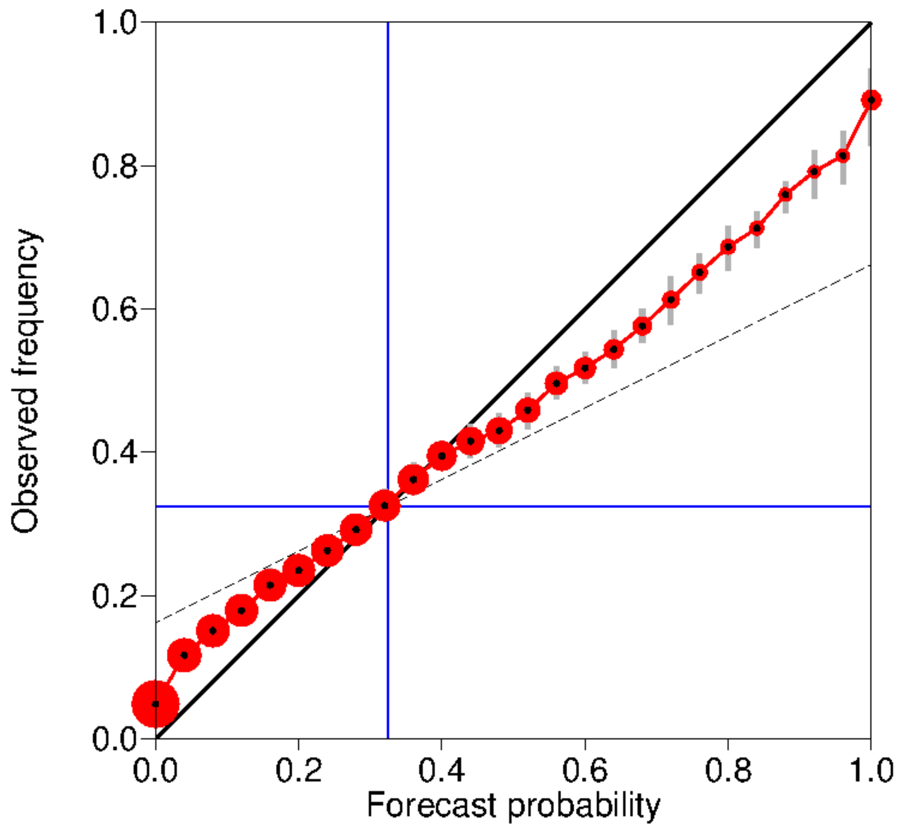



Anomaly Correlation Coefficient for 0001 with 25 ensemble members
Near-surface air temperature
Hindcast period 1981-2016 with start in September average over months 2 to 4
Black dots for values significantly different from zero with 95% confidence (1000 samples)
-1 -0.9 -0.8 -0.7 -0.6 -0.4 -0.2 0.2 0.4 0.6 0.7 0.8 0.9 1

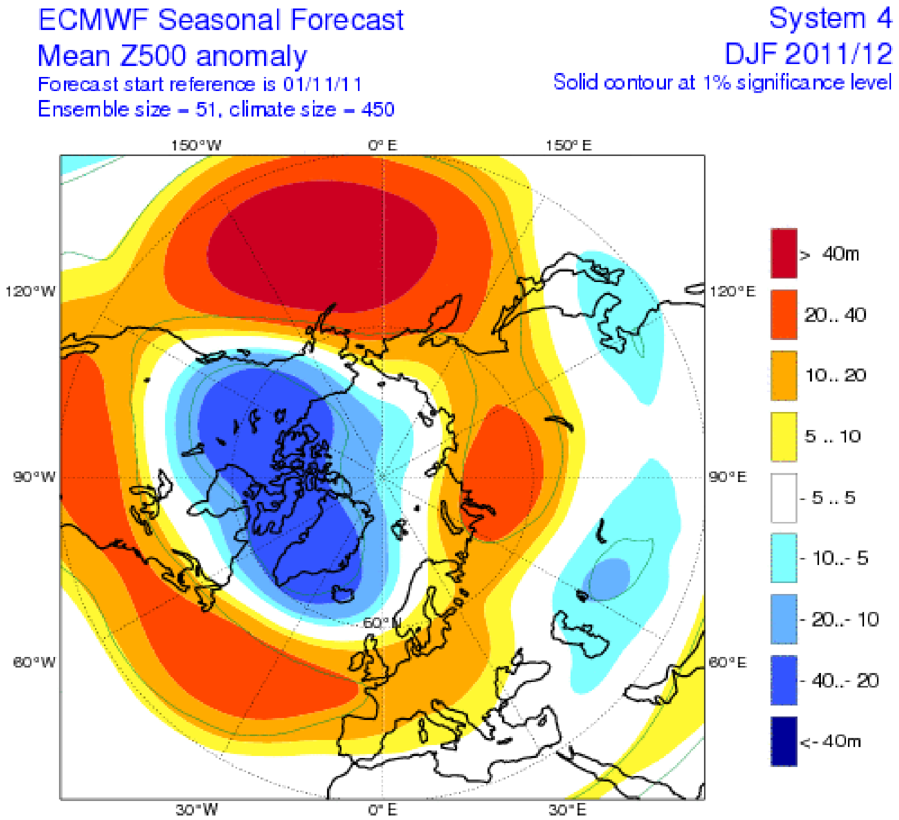
ROC Skill Score for ORecmfEX0001SY05M1 with 25 ensemble members and 26 bins
Near-surface air temperature anomalies above the upper tercile
Hindcast period 1981-2016 with start in September and averaging period 2 to 4
Threshold estimated with a kernel method for the PDF

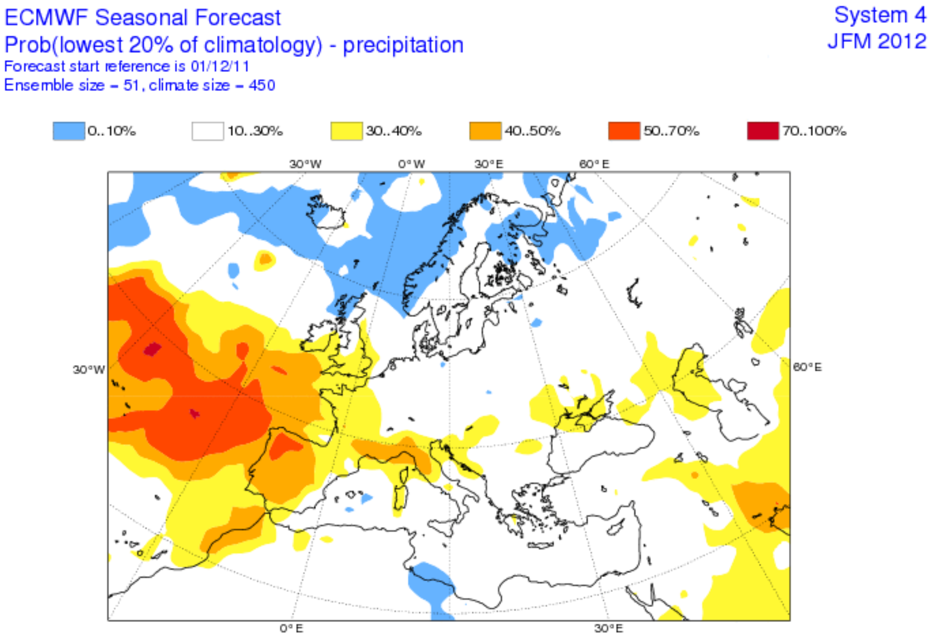


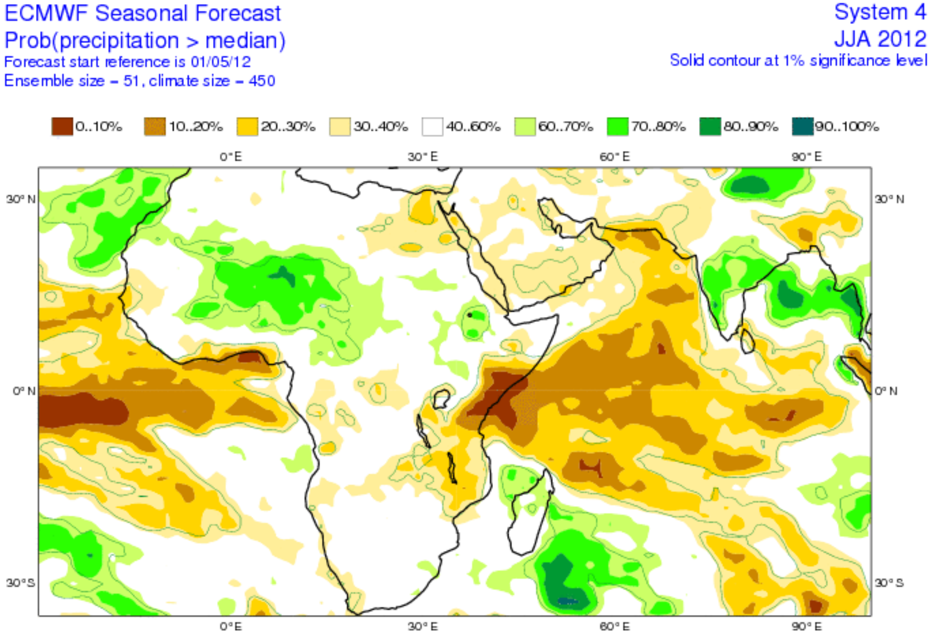
Reliability diagram for 0001 with 25 ensemble members
Near-surface air temperature anomalies above the upper tercile
Accumulated over global (land and sea points)
Hindcast period 1981-2016 with start in September average over months 2 to 4
Skill scores and 95% conf. intervals (1000 samples)
Brier skill score: 0.175 (0.134, 0.215)
Reliability skill score: 0.982 (0.976, 0.987)
Resolution skill score: 0.193 (0.155, 0.231)



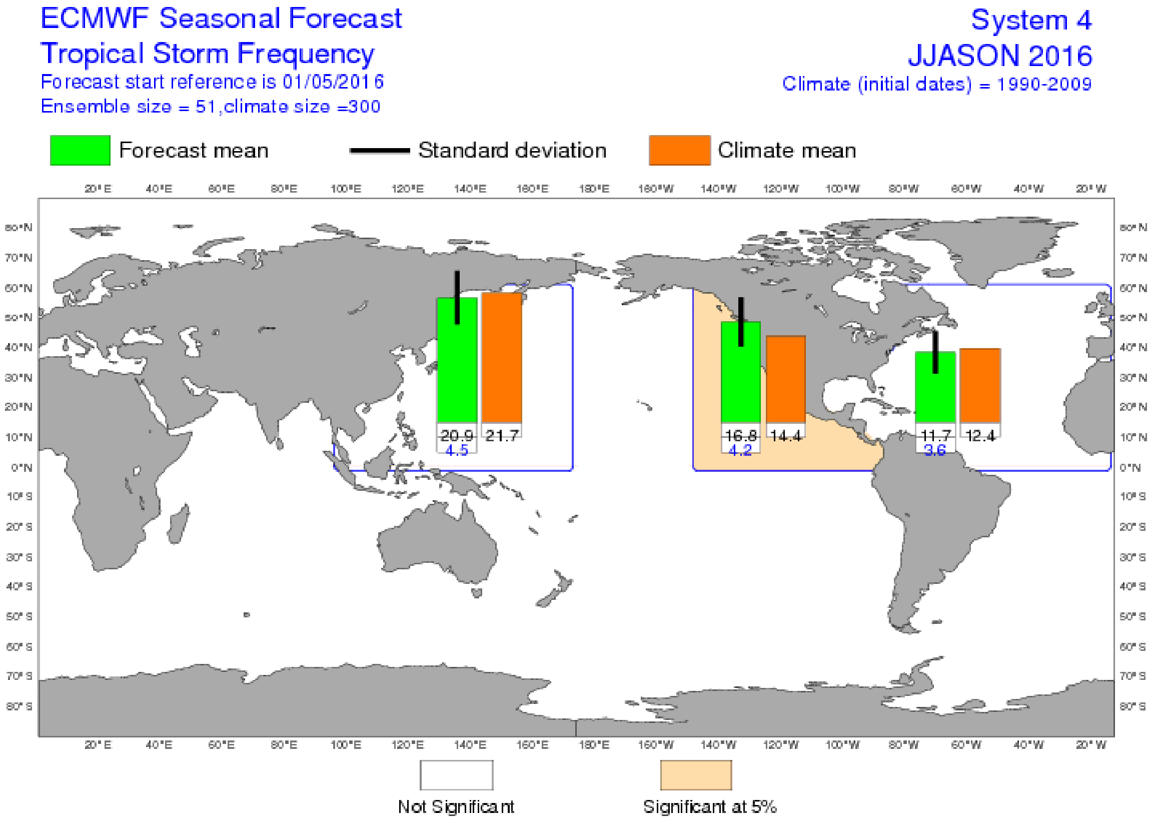
Other operational plot examples

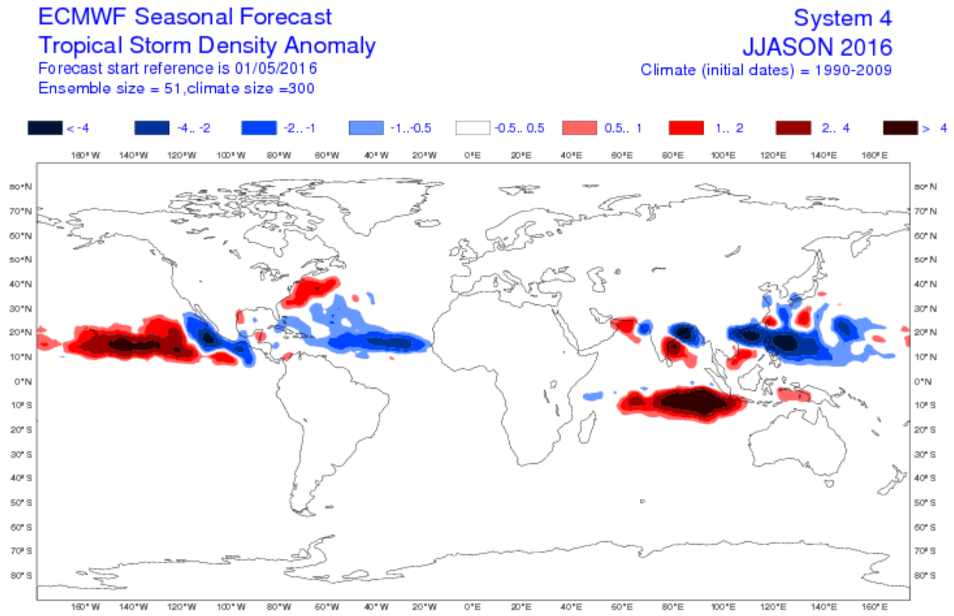


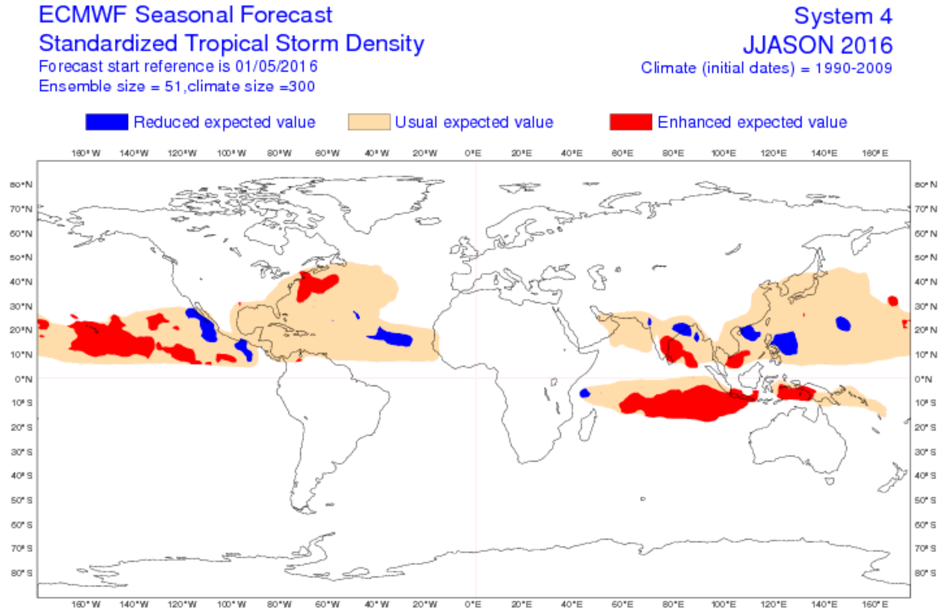




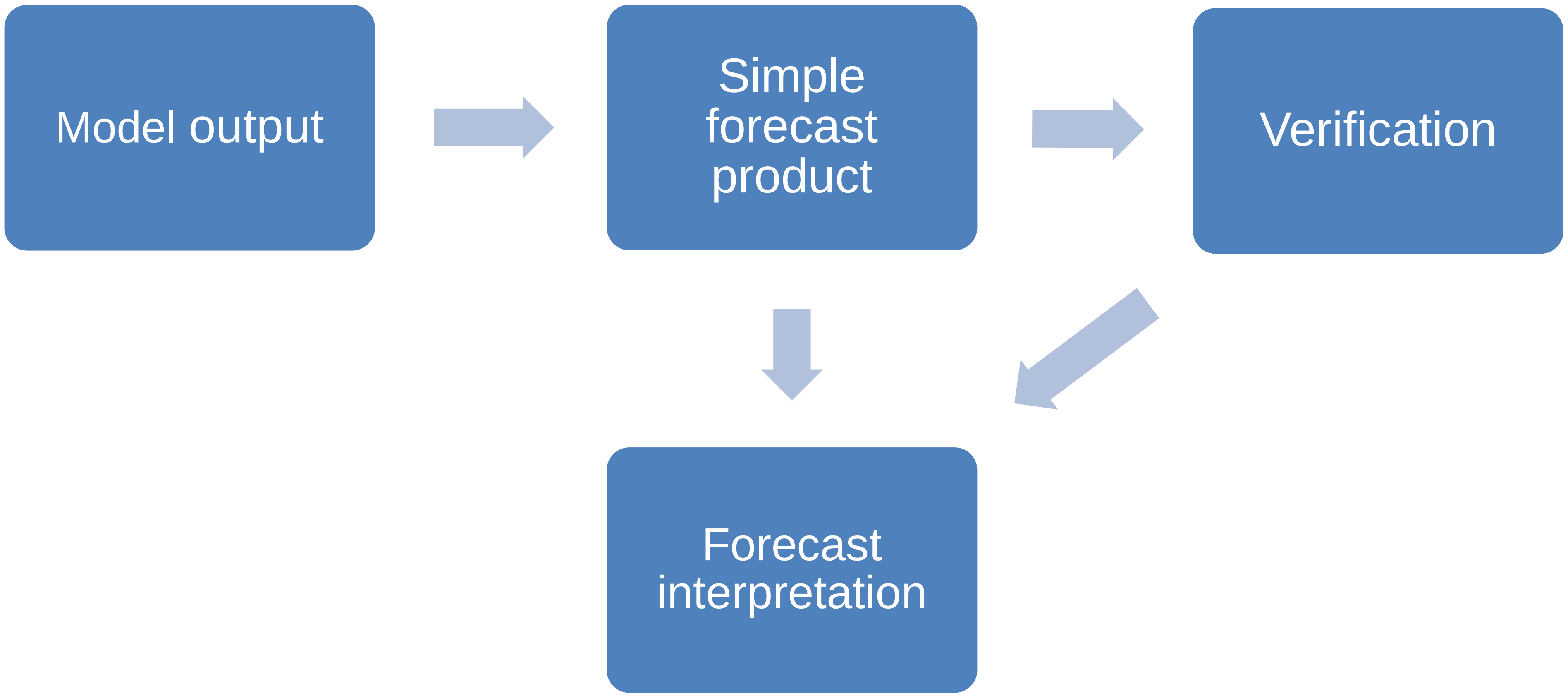
Tropical storm forecasts



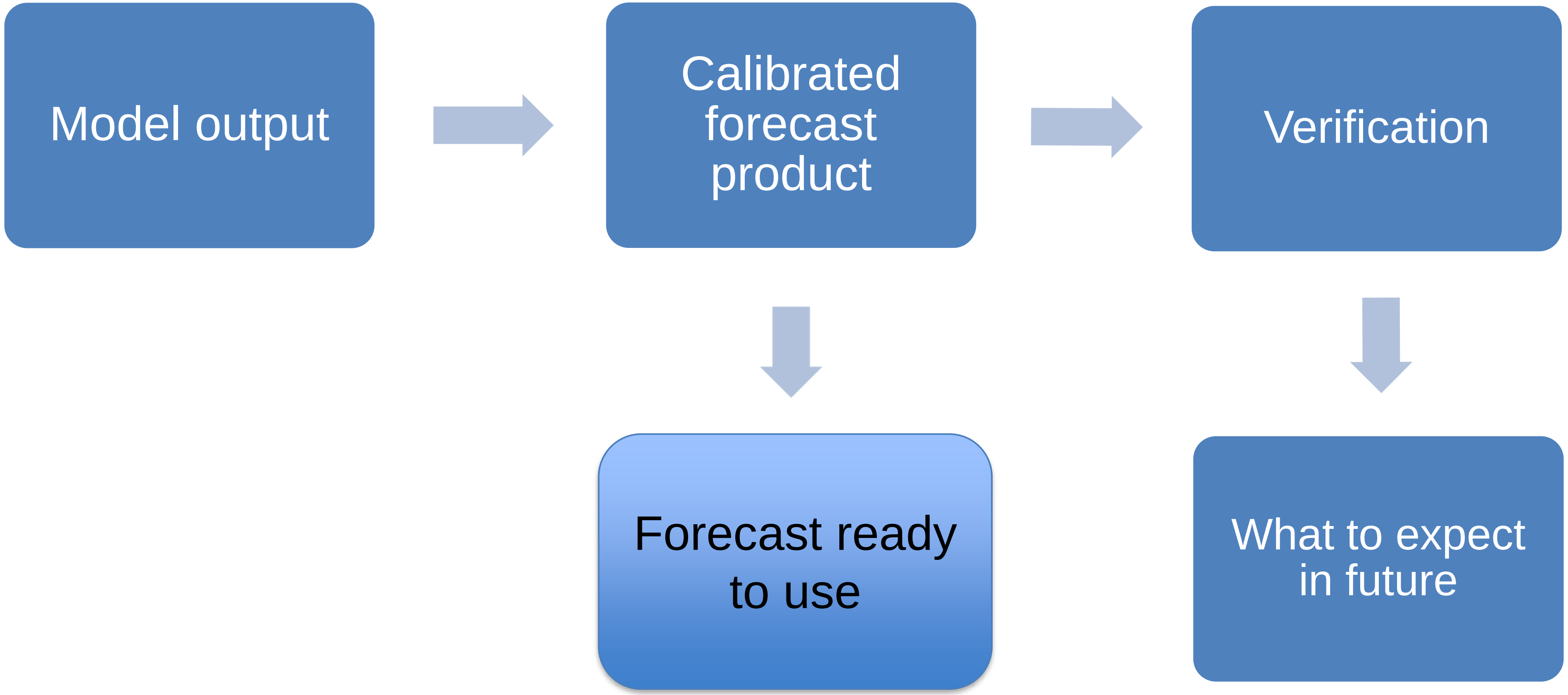




4. Forecast process I: How it is often done (including at ECMWF)

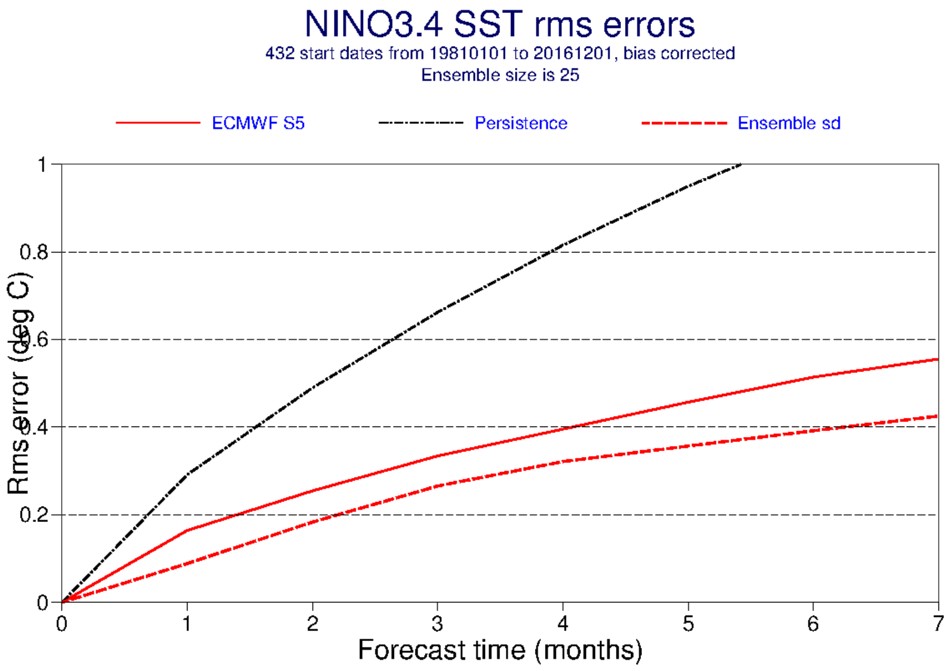


Forecast process II: How it should be done



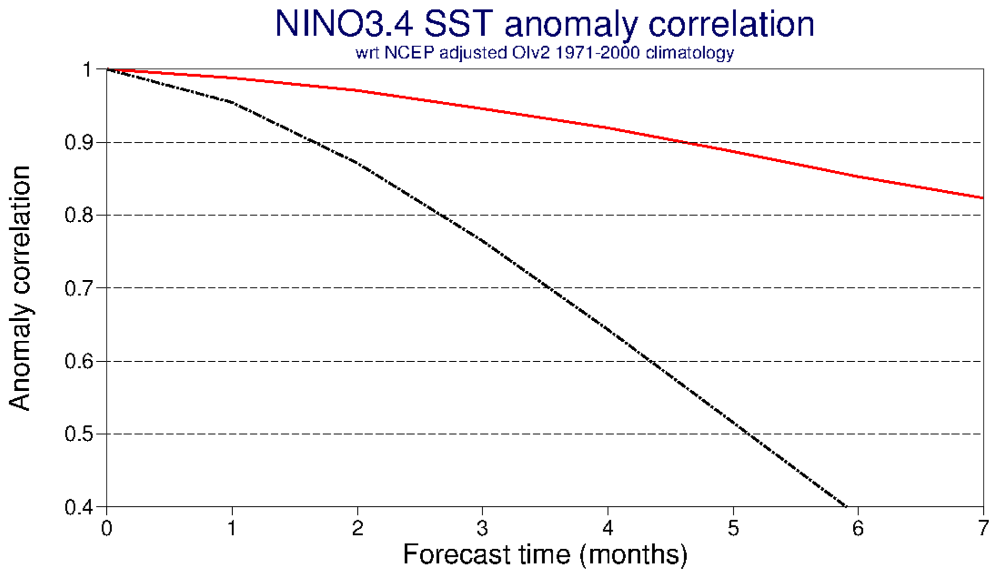
SST forecast performance

Actual r.m.s. errors > model estimate of "perfect model" errors

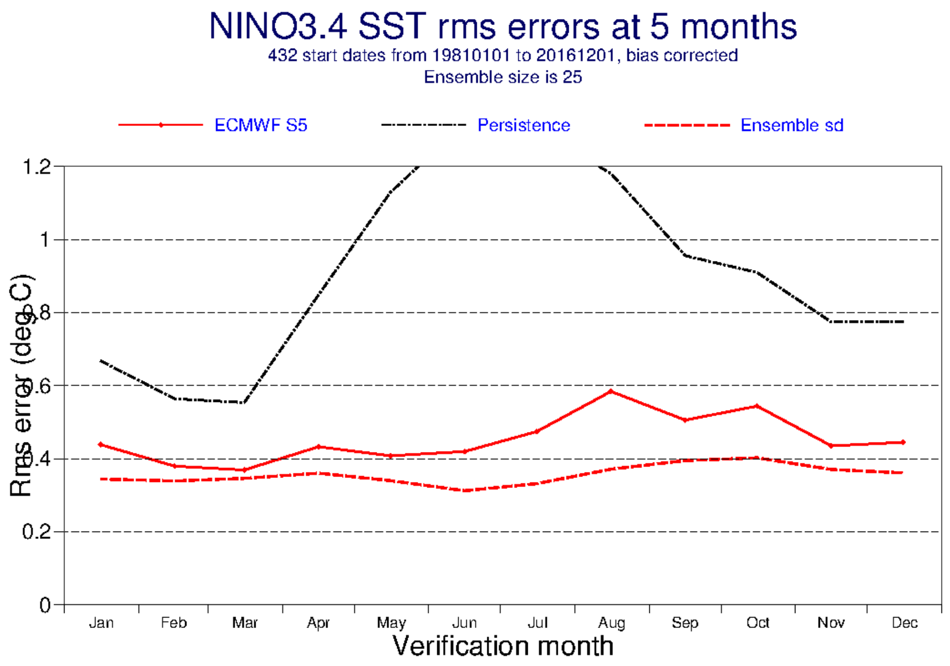


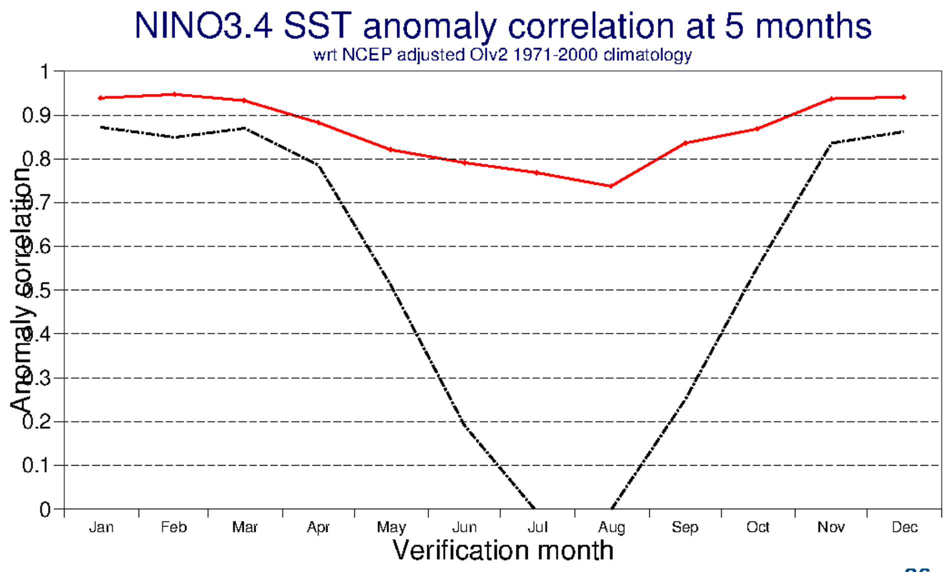
NOTE: Stochastic physics gives increased spread to Nino SSTs, due to its representation of low-frequency model error.

This gives better probabilistic scores, but means the ensemble spread is not a predictability limit: if in future systems we reduce the model error, we can reduce the amplitude of the stochastic "noise" to match, and the ensemble spread will reduce.



Seasonal dependence

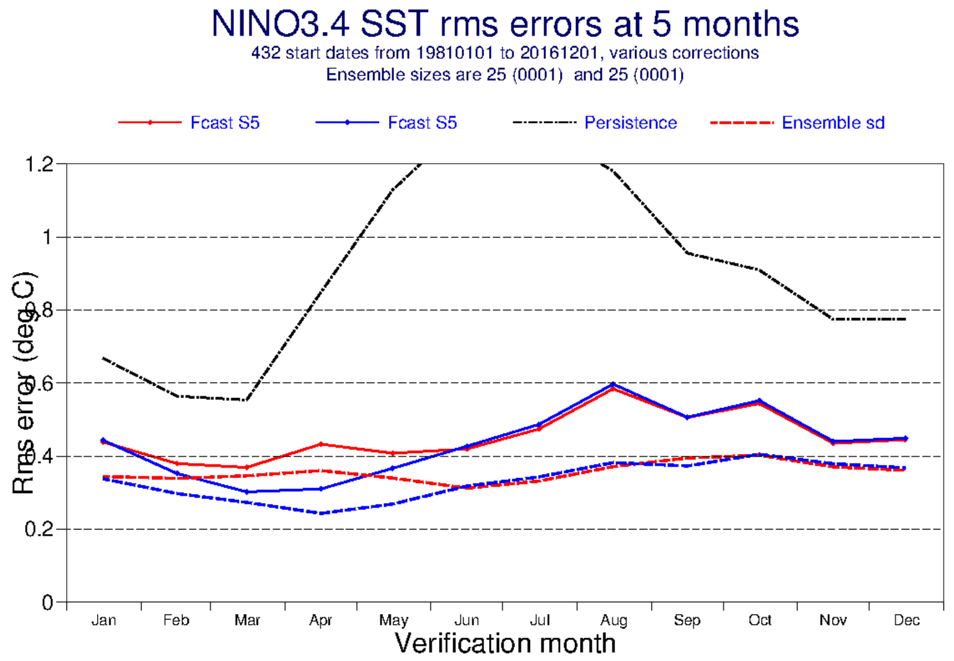


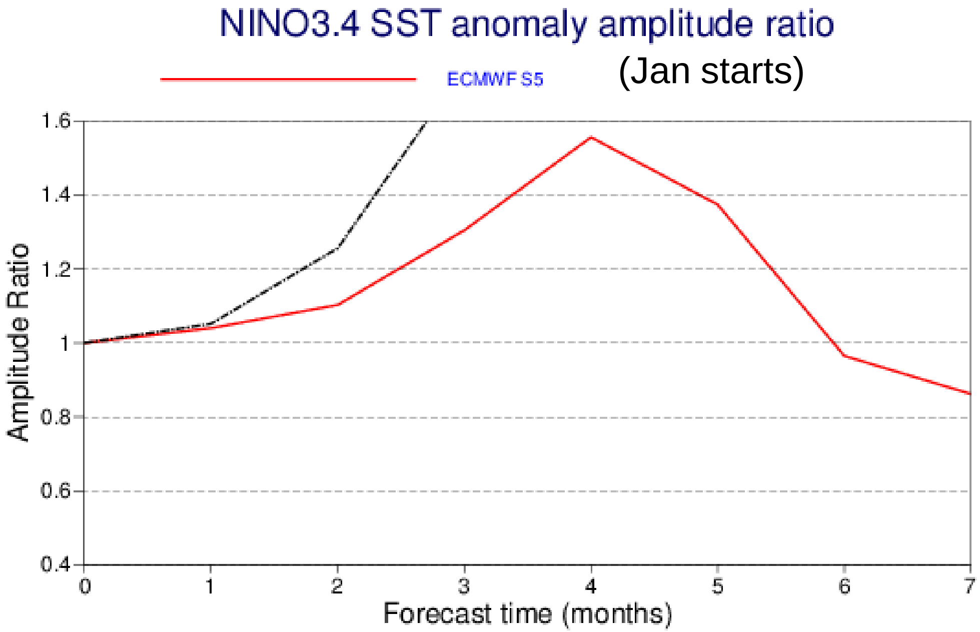


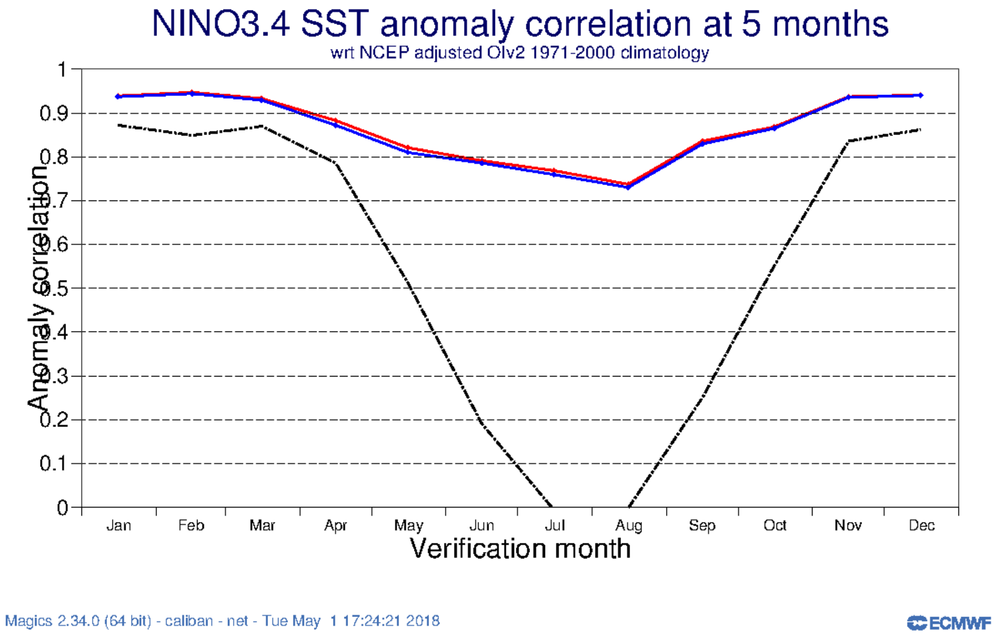
Variance adjustment

This very simple calibration scales the forecast climatological variance to match the observed climatological variance. The scaling is seasonally dependent. This calibration can substantially improve forecast products (and their verification scores). This calibration was used for our previous system, but was turned off in SEAS5.

SEAS5 verification includes the amplitude ratio, which should be used *a posteriori* to interpret the Nino plumes. This is important for forecasts of March, April and May.



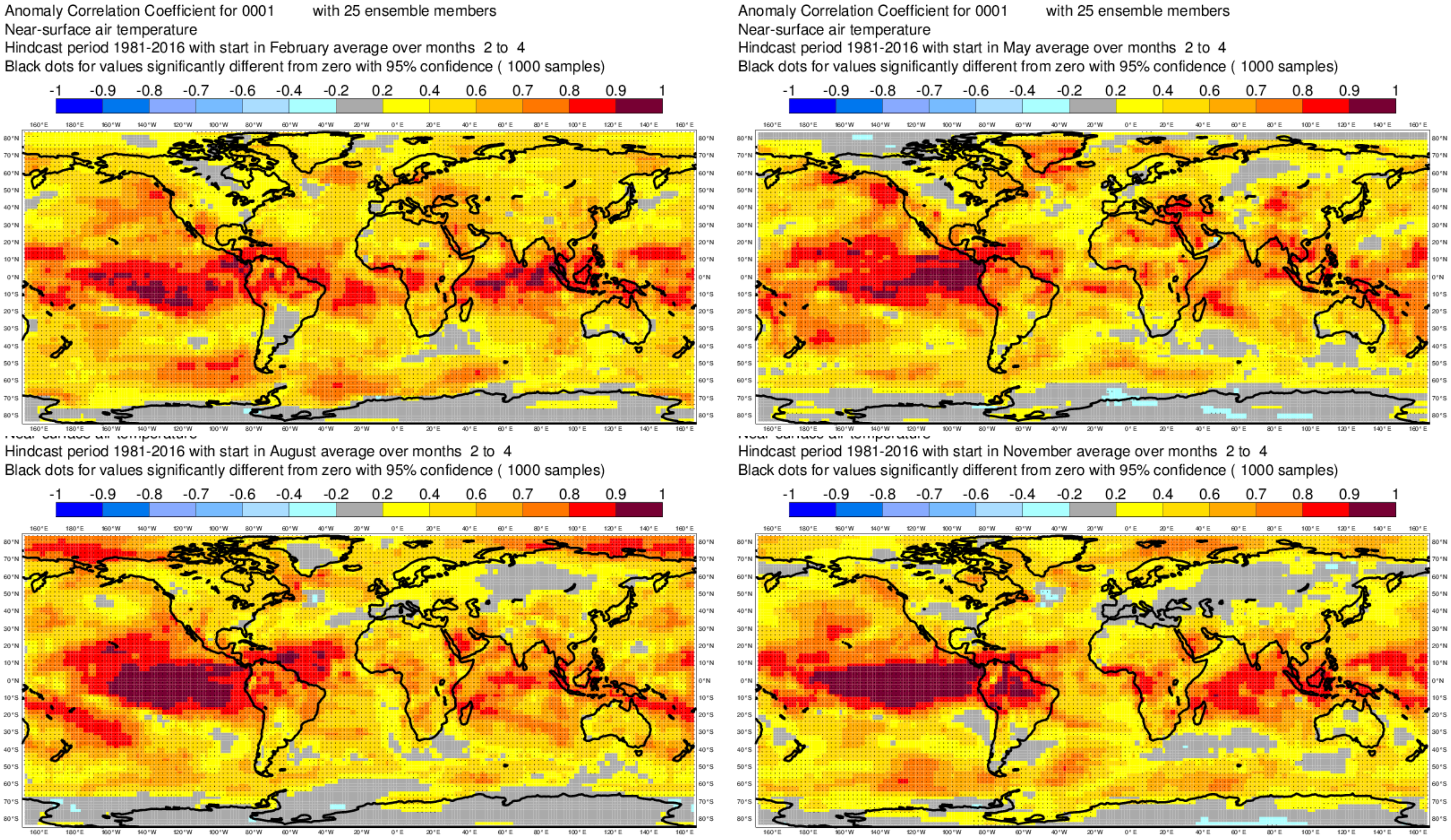




Seasonal dependence of surface parameters

MAM

JJA



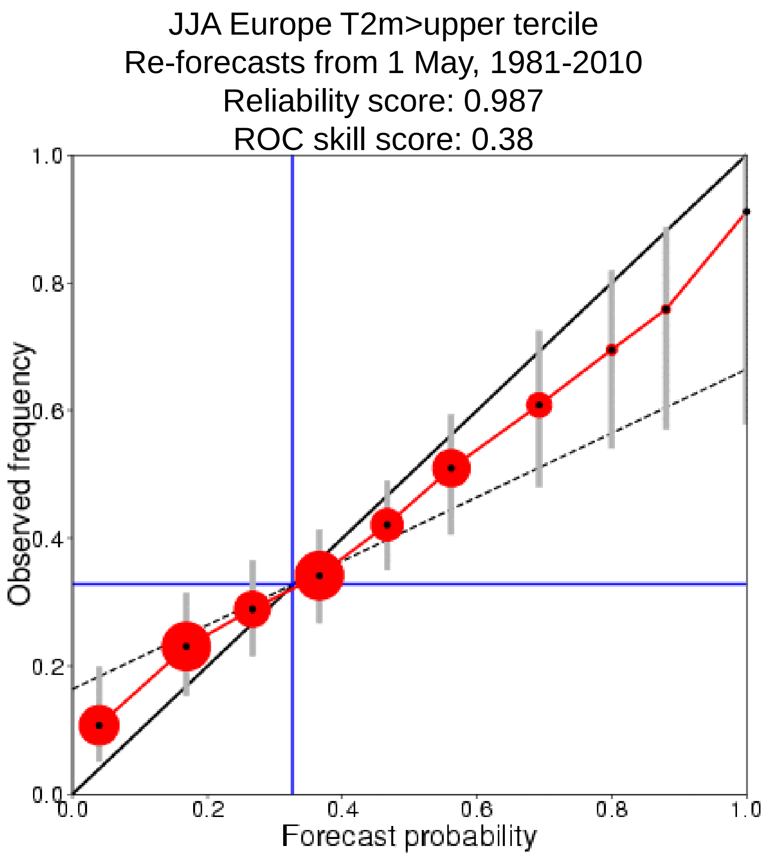
SON

DJF

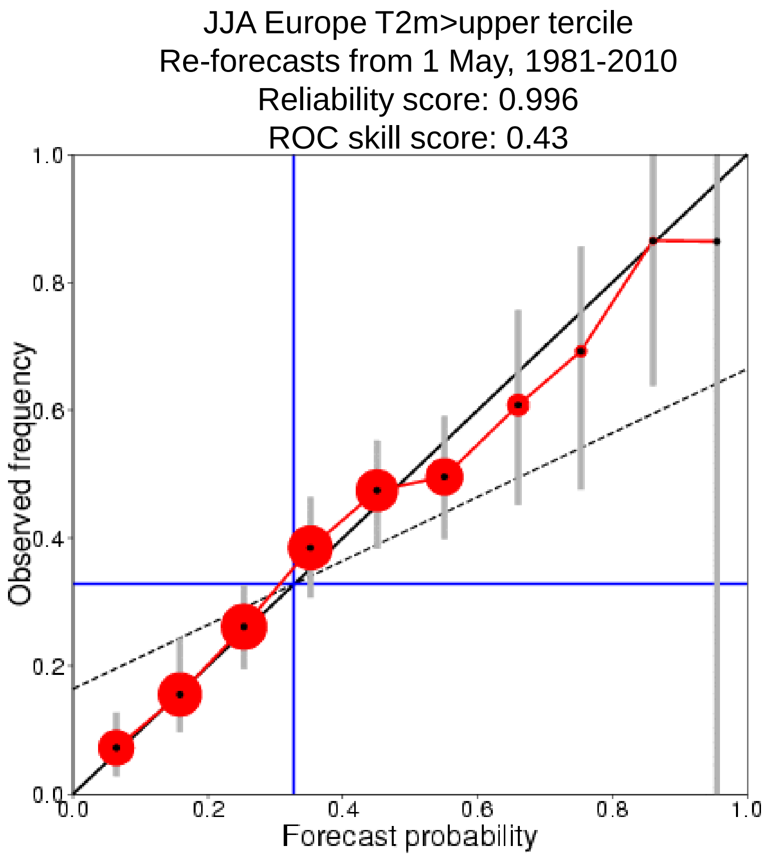
Role of ensemble size: Scores for Europe in JJA

(System 4)

15 members

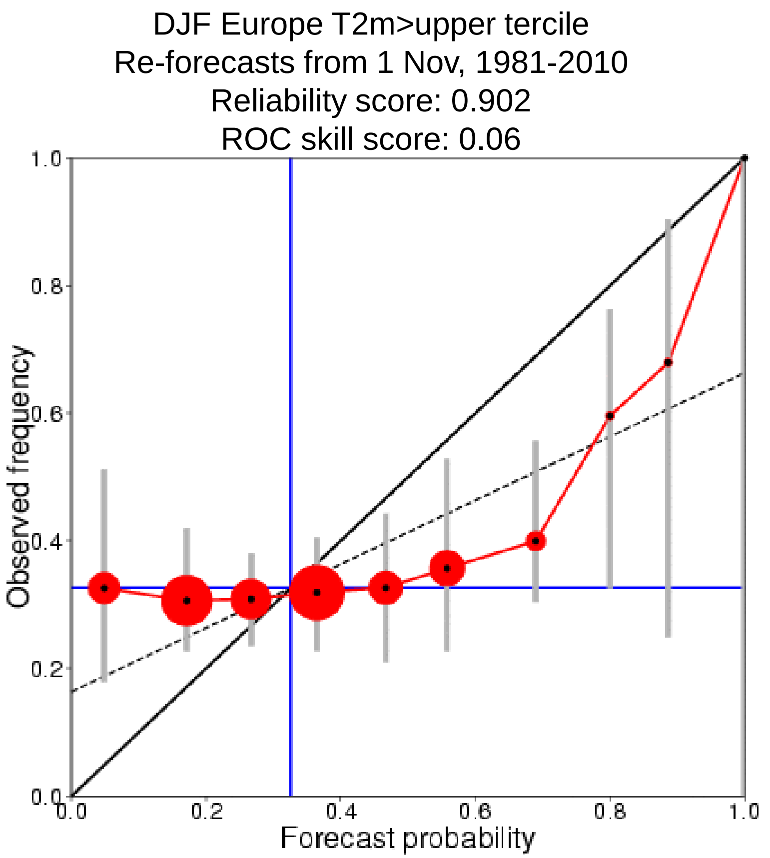


51 members

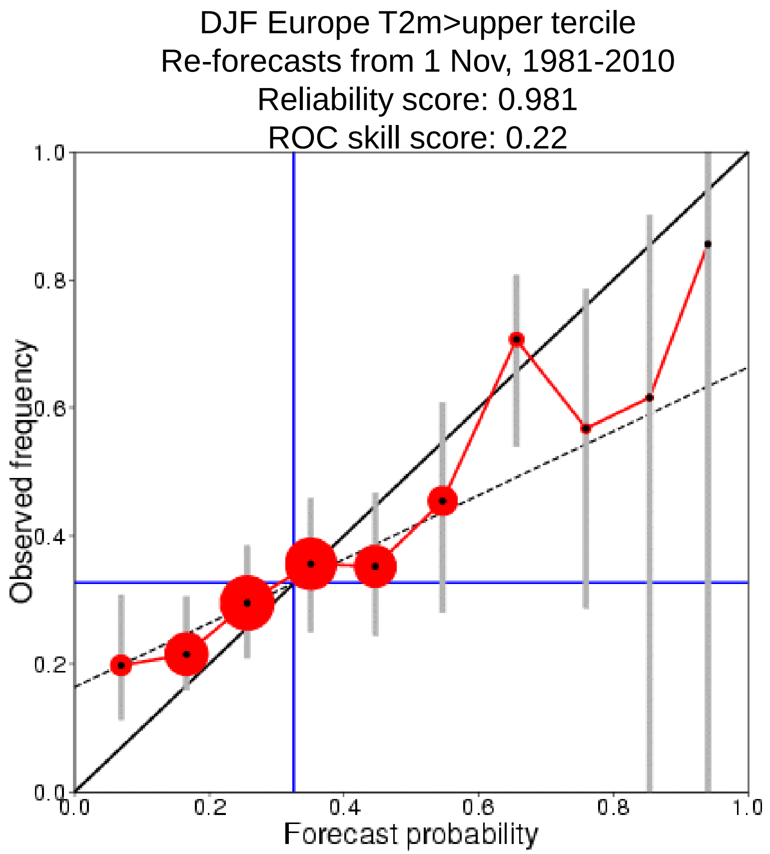


Scores for Europe: DJF

15 members



51 members



Overall assessment: how good are the forecasts?

The best way to answer this question is to browse the various skill scores and maps made available on the ECMWF (or other provider) website. But a highly simplified summary is:

- Skill (ACC, BSS, ROC, …) relative to climate is typically moderate to high in the tropics, moderately low to sometimes very low in mid-latitudes.
- Reliability is on average moderately good; large ensemble sizes are needed to measure this in low-predictability areas.
- Even with large ensemble sizes, the limited number of years means that skill assessments have large uncertainties in mid-latitude regions.
- We can average skill over many gridpoints, seasons etc. to try to reduce uncertainty, but we inevitably trade spatial resolution to gain a bit more accuracy.

- As we are about to see, there are indications that in some cases the forecast spread is too large.

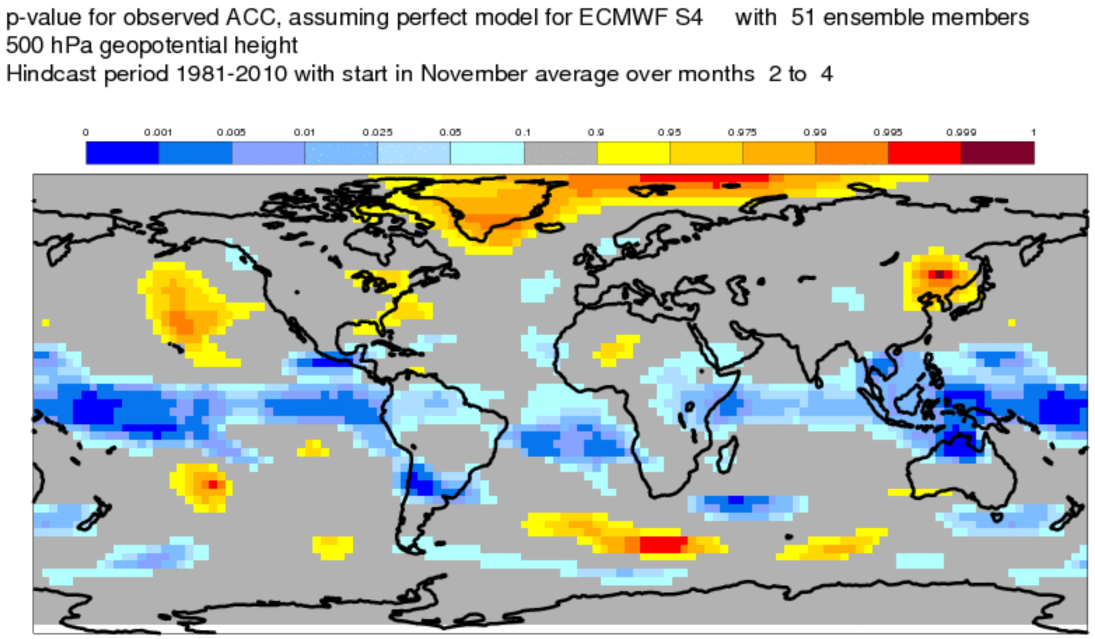
5. Future Challenges and prospects

- NH winter predictability

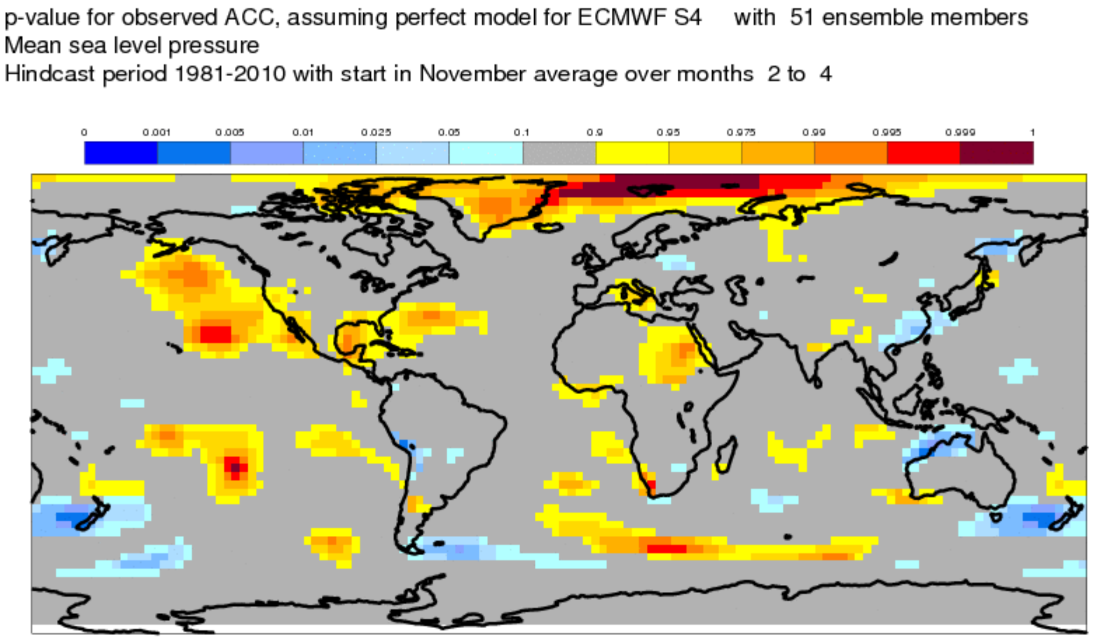
- QBO teleconnections

Predictive skill vs. Predictability limit

DJF Z500



DJF MSLP



Indistinguishable from perfect
Worse than perfect
Better than perfect

Predictability of the Arctic Oscillation

Predictability can be under-estimated if we miss or under-represent important processes

Scope for improvement

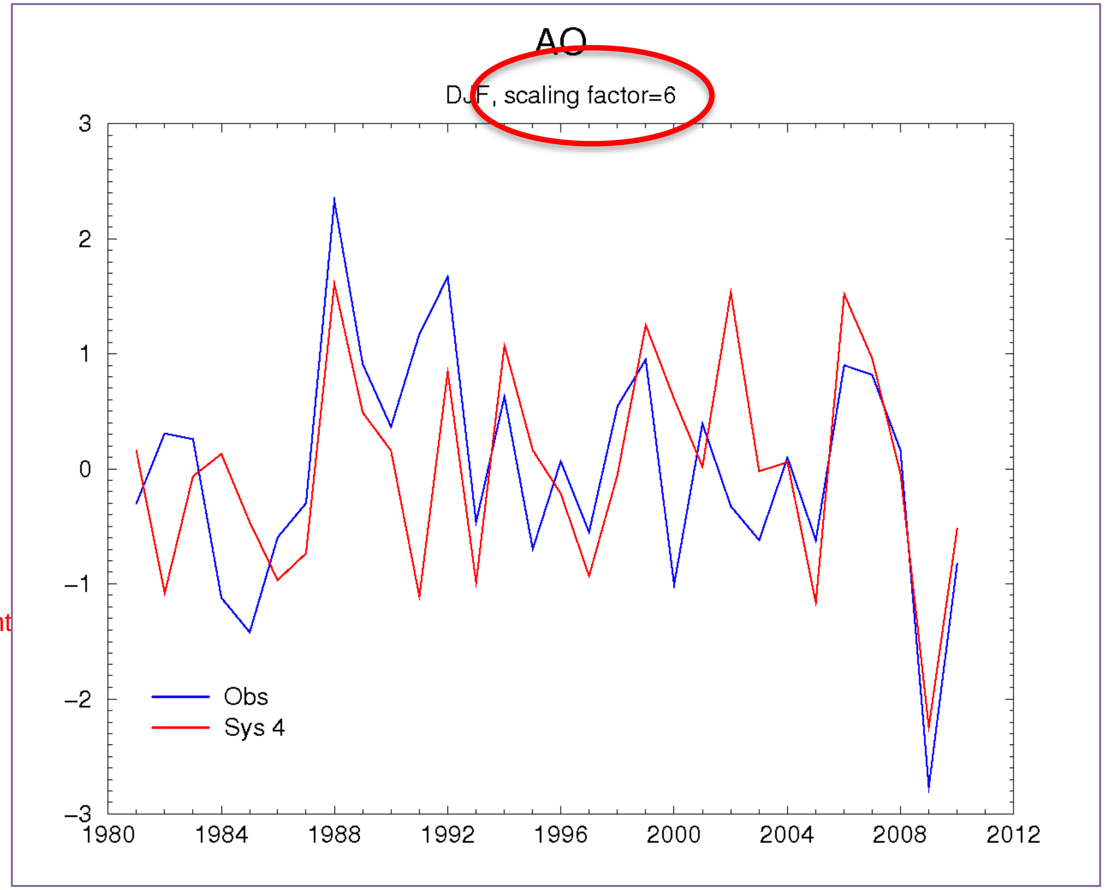
```
Unbiased variance estimates: Obs/Tot/Int/Ext:      1.0000      0.8390      0.8316

Model/obs stddev ratio:   0.9159
Model/obs stddev ratio interval:       0.693   1.129
Bootstrap over nens, pval for ratio=1:  0.7960

==========================================
SNR actual                      :   0.0941
SNR jackknife over nens         :   0.0202  0.1029  0.1857
==========================================

==========================================
ACC actual                      :   0.6085
ACC basic bootstrap over nens   :   0.5568  0.7121  0.8144
ACC basic bootstrap over nyears:   0.2052  0.6069  0.8326
==========================================

ACP from internal sampling: -0.2947  0.0583  0.4010
Mean ACC for nens-1:   0.6049
p val of measured acc if model perfect:   0.9996
```

← model variability consistent

← 95% interval due to ensemble size

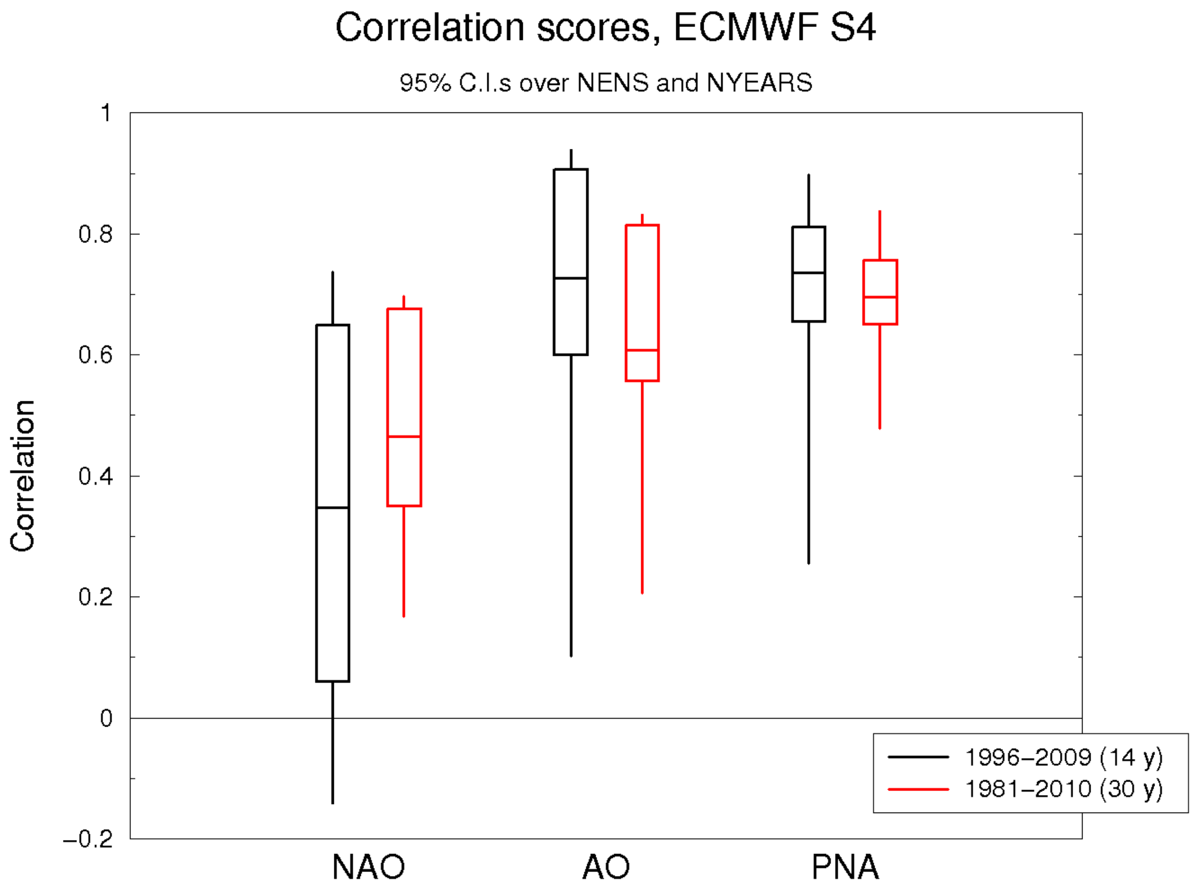
← bigger uncertainty range here

← only a 0.0004 chance we could get this correlation



Correlation (30y) =0.608

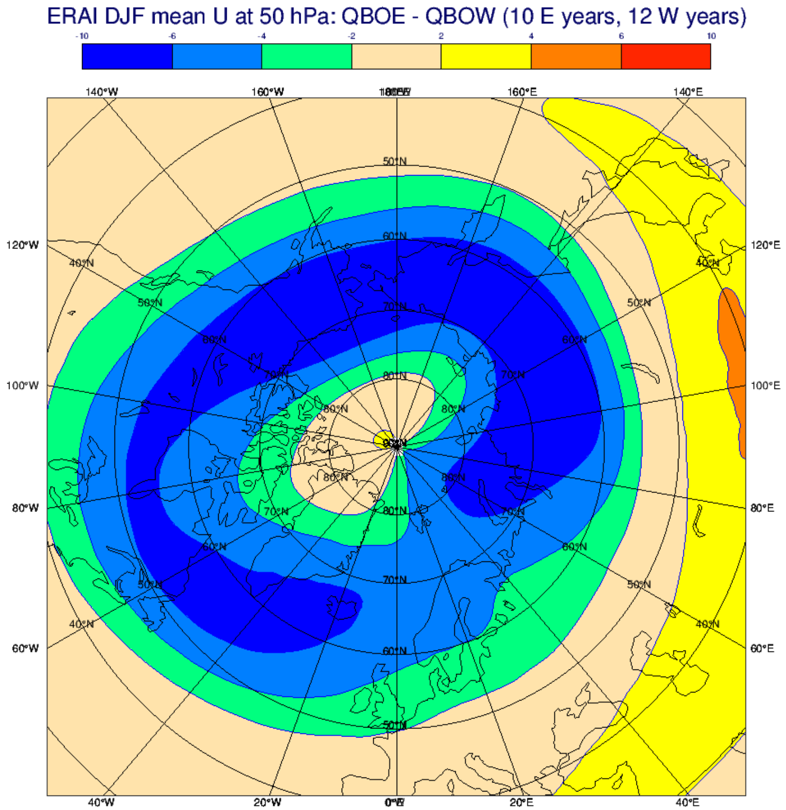
Challenge: sampling errors are large!



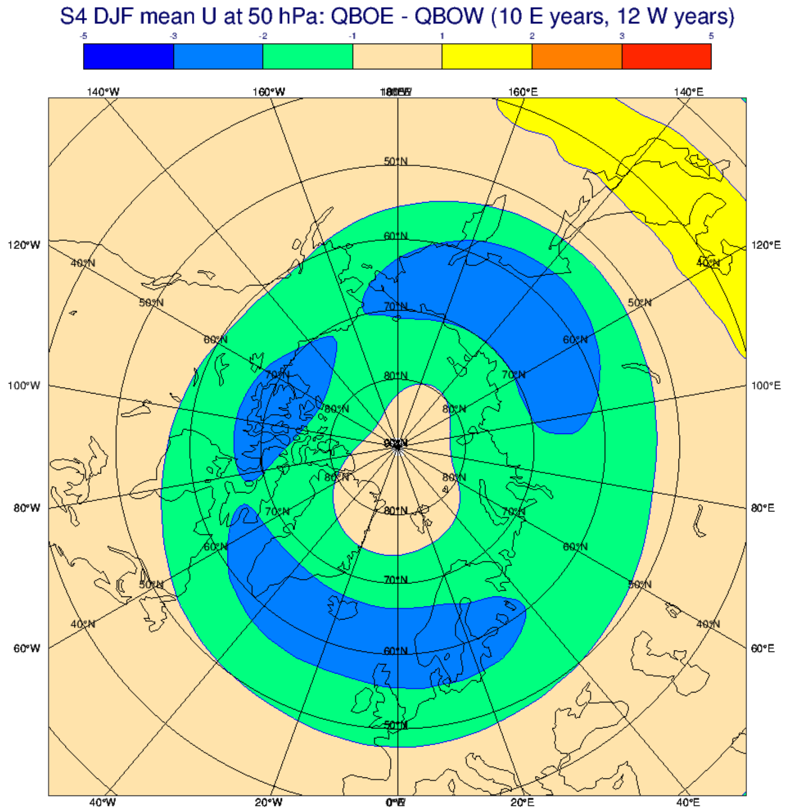
Box = 95% interval, bootstrapping on ensemble size

Whiskers = 95% interval, bootstrapping on years included

QBO teleconnections – NH winter polar vortex (50 hPa)

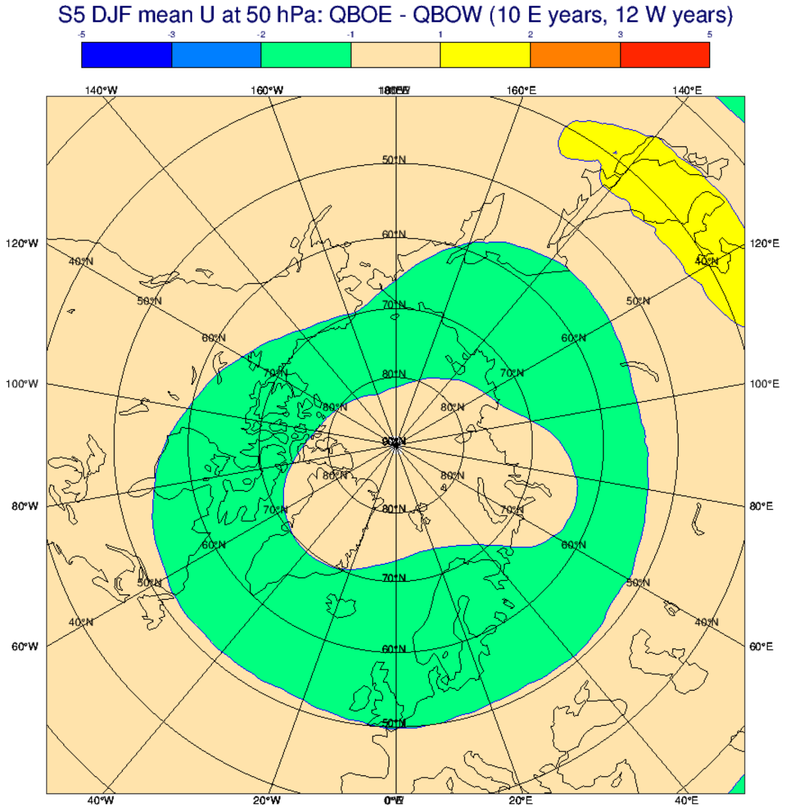


System 4



Factor 3 weaker than obs

SEAS5



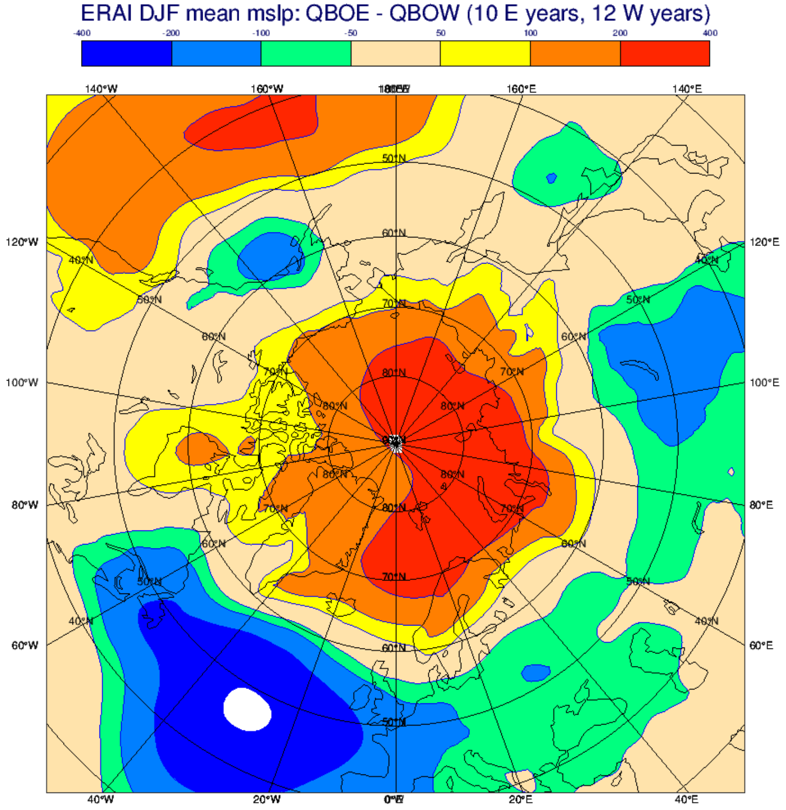
Factor 5 weaker than obs

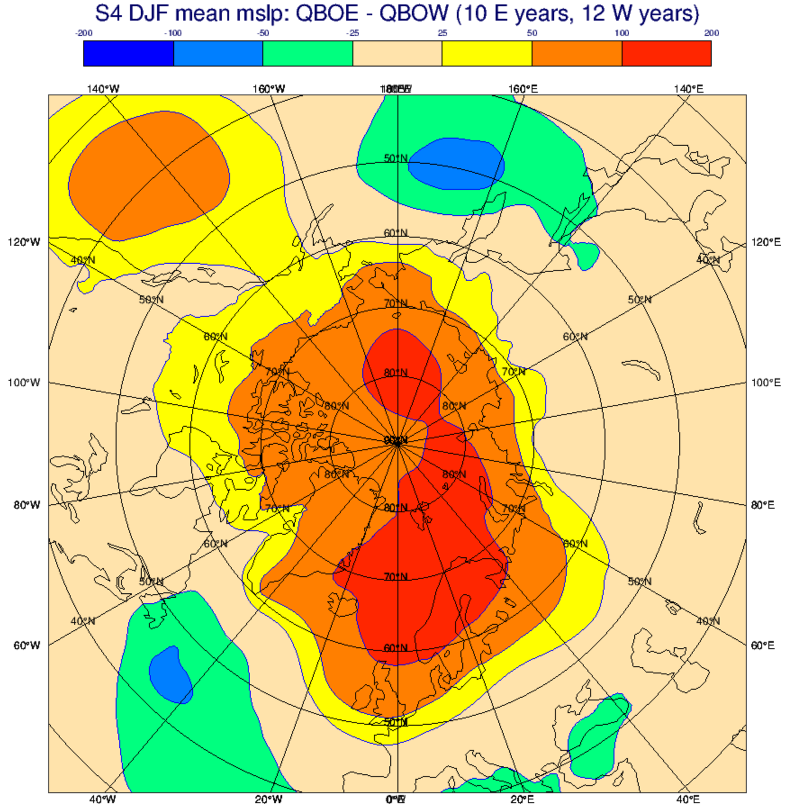
QBO composite years for 1981-2005, following Boer and Hamilton (2008). Contour interval is 2 m/s for ERAI, 1 m/s for models. Model composites based on 25 member ensemble.

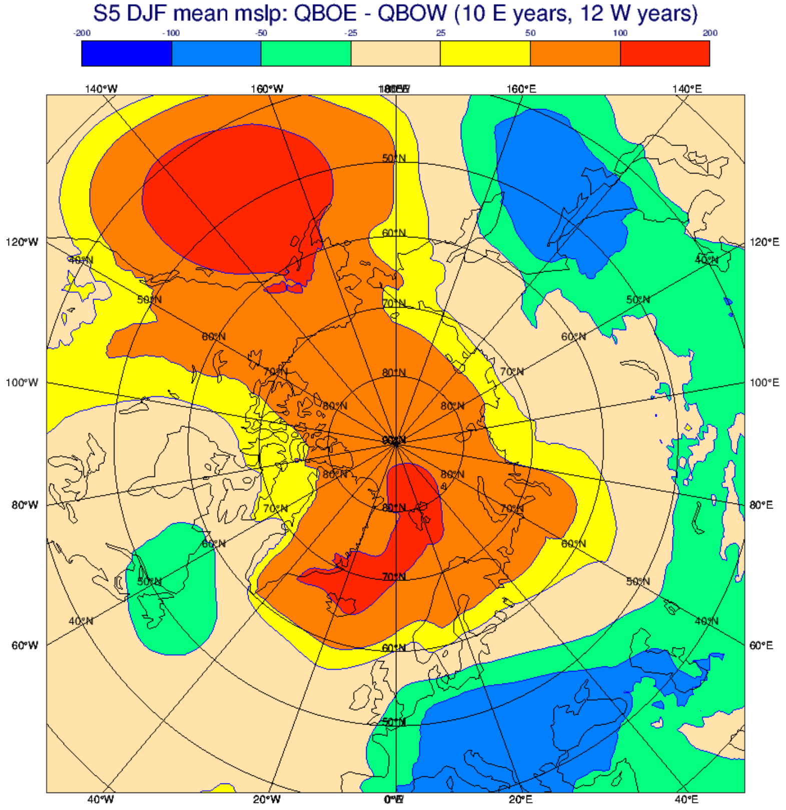
QBO teleconnections – NH winter MSLP

System 4

SEAS5







QBO composite years for 1981-2005, following Boer and Hamilton (2008). Contour interval is 0.5 hPa for ERAI, 0.25 hPa for model. Model composites based on 25 member ensemble.

6. Final perspectives

- Seasonal prediction skill is variable and reliability is moderate, but both skill and reliability are at a level where many applications are possible.
- Care is needed to ensure forecast information is properly interpreted and used sensibly

- Forecasting models are fairly realistic in many ways, but remaining errors are enough to substantially impact forecast skill and reliability, even after calibration
- Creating consistent initial conditions for past and present is a challenge, due in particular to the lack of observational data in the past. Observing systems are better now, but still need some improvements.
- Limited predictability and limited past data prevent us being sure about the skill levels of today's forecast systems, and calibration is therefore subject to uncertainty.

- Although multi-model ensembles are helpful, they only partially span the space of model errors.
- In the end, the only way to achieve high reliability is to build trustworthy models

References and further reading

SEAS5 forecasts on www.ecmwf.int/en/forecasts/charts and https://climate.copernicus.eu/seasonal-forecasts

ECMWF Seasonal Forecast User Guide

SPECS fact sheets http://www.specs-fp7.eu/Fact%20sheets on seasonal forecasting

Boer, G.J. and K. Hamilton, 2008: QBO influence on extratropical predictive skill. Clim. Dyn. 31:987–1000. doi: 10.1007/s00382-008-0379-5

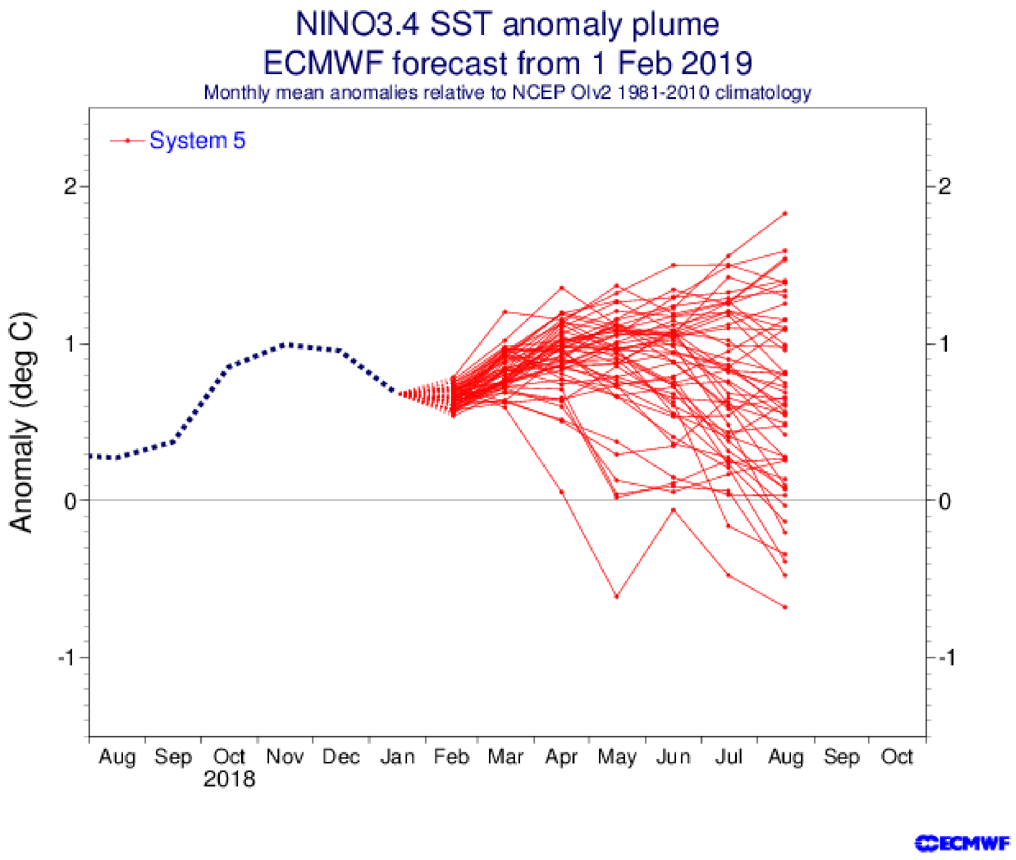
Molteni, F., Stockdale, T., Balmaseda, M., Balsamo, G., Buizza, R., Ferranti, L., Magnusson, L., Mogensen, K., Palmer, T., and Vitart, F, 2011: The new ECMWF seasonal forecast system (System 4), ECMWF Tech. Memo 656., DOI:10.21957/4nery093i

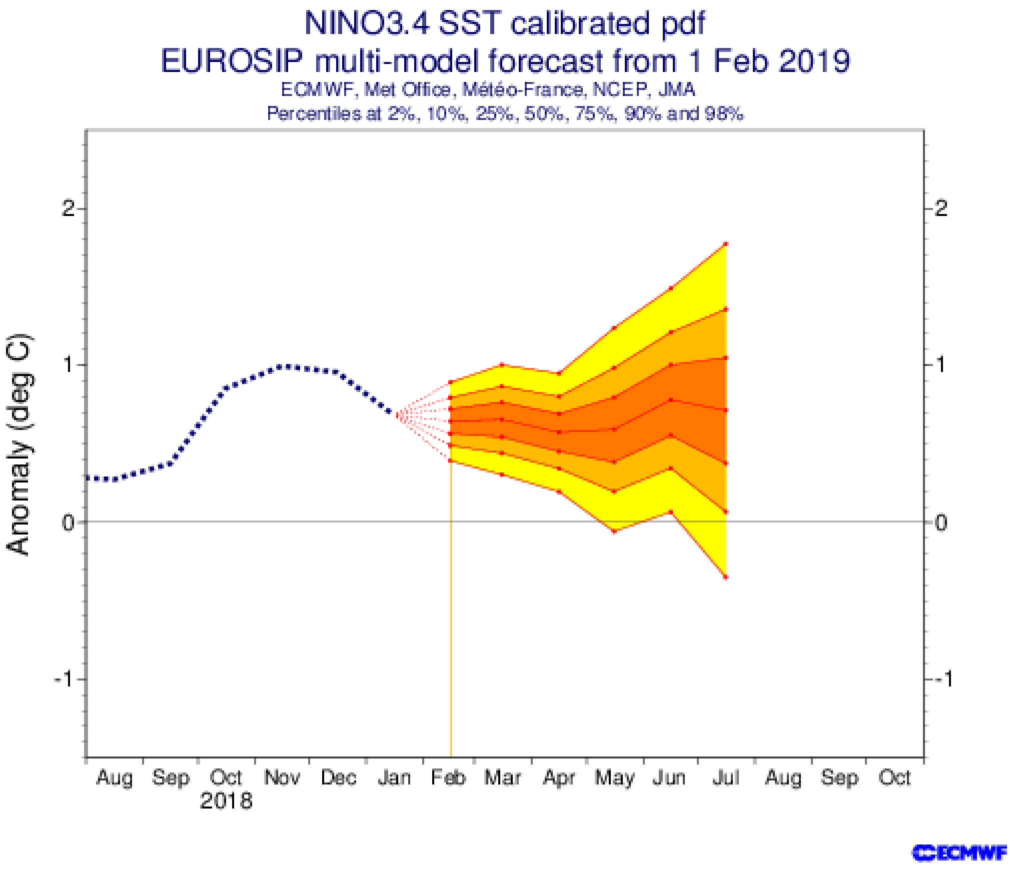
Stockdale T.N., F. Molteni and L. Ferranti, 2015: Atmospheric initial conditions and the predictability of the Arctic Oscillation. Geophys. Res. Lett. doi: 10.1002/2014GL062681

Johnson, S. J., Stockdale, T. N., Ferranti, L., Balmaseda, M. A., Molteni, F., Magnusson, L., Tietsche, S., Decremer, D., Weisheimer, A., Balsamo, G., Keeley, S., Mogensen, K., Zuo, H., and Monge-Sanz, B., 2018: SEAS5: The new ECMWF seasonal forecast system, Geosci. Model Dev. DOI: 10.5194/gmd-2018-228

Stockdale, T., Alonso-Balmaseda, M, Johnson, S, Ferranti, L, Molteni, F, Magnusson, L, Tietsche, S, Vitart, F, Decremer, D, Weisheimer, A, Roberts, CD, Balsamo, G, Keeley, S, Mogensen, K, Zuo, H, Mayer, M, and Monge-Sanz, BM, 2018: SEAS5 and the future evolution of the long-range forecast system. ECMWF Tech Memo 835, DOI: 10.21957/z3e92di7y

Latest ENSO forecast





Predicted MAM 2018 (left) and 2019 (right)