Verification of temperature and precipitation trends in climate models.

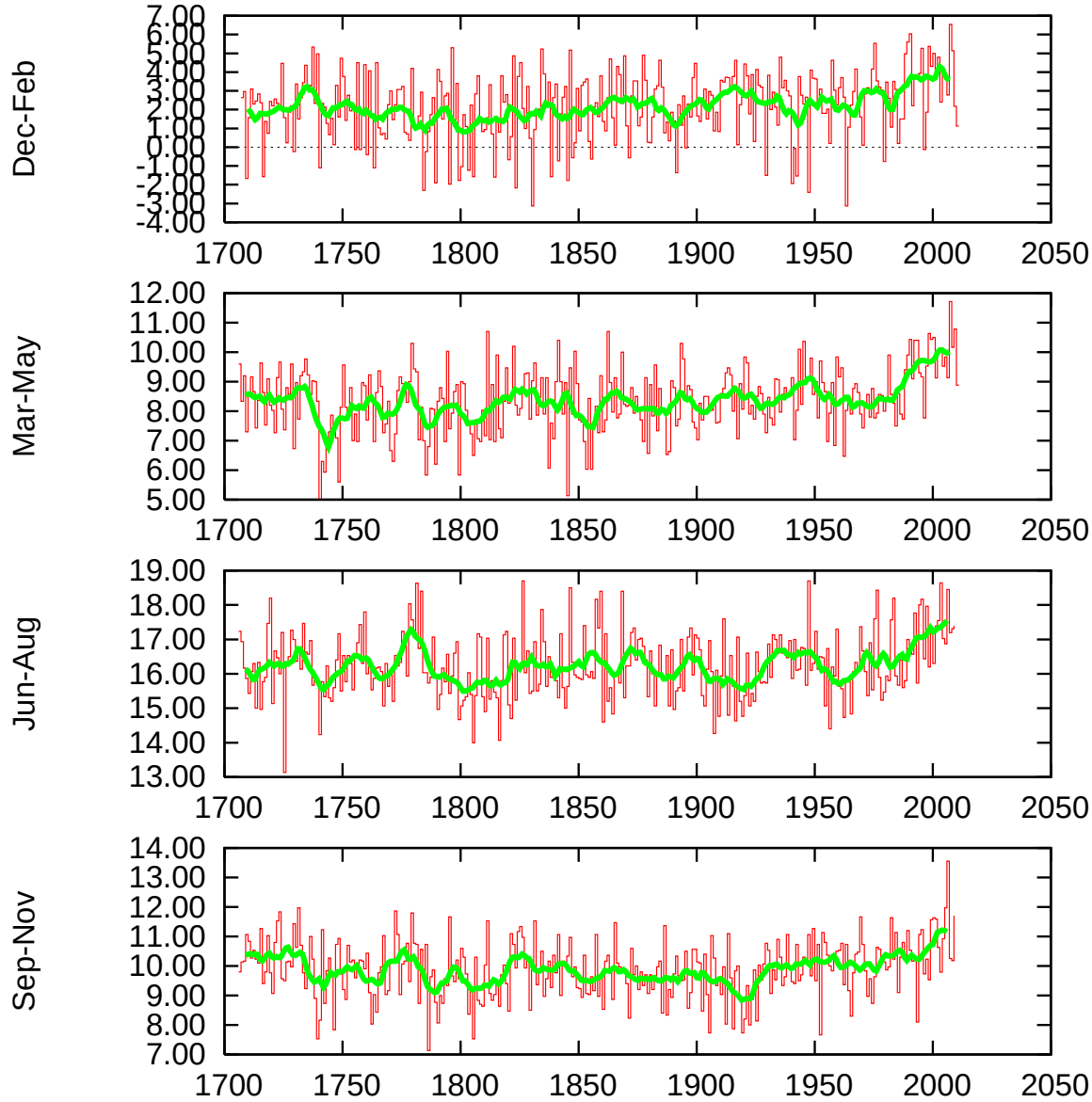
Geert Jan van Oldenborgh, KNMI
many co-authors

- Introduction
- Uninitialised climate simulations
  - Observed temperature trends
  - Climate model temperature trends
  - Statistical analysis of the discrepancies
  - Physical causes of the discrepancies
- Initialised climate simulations
  - Global temperature
  - Temperature fields
  - AMO index
  - Precipitation fields

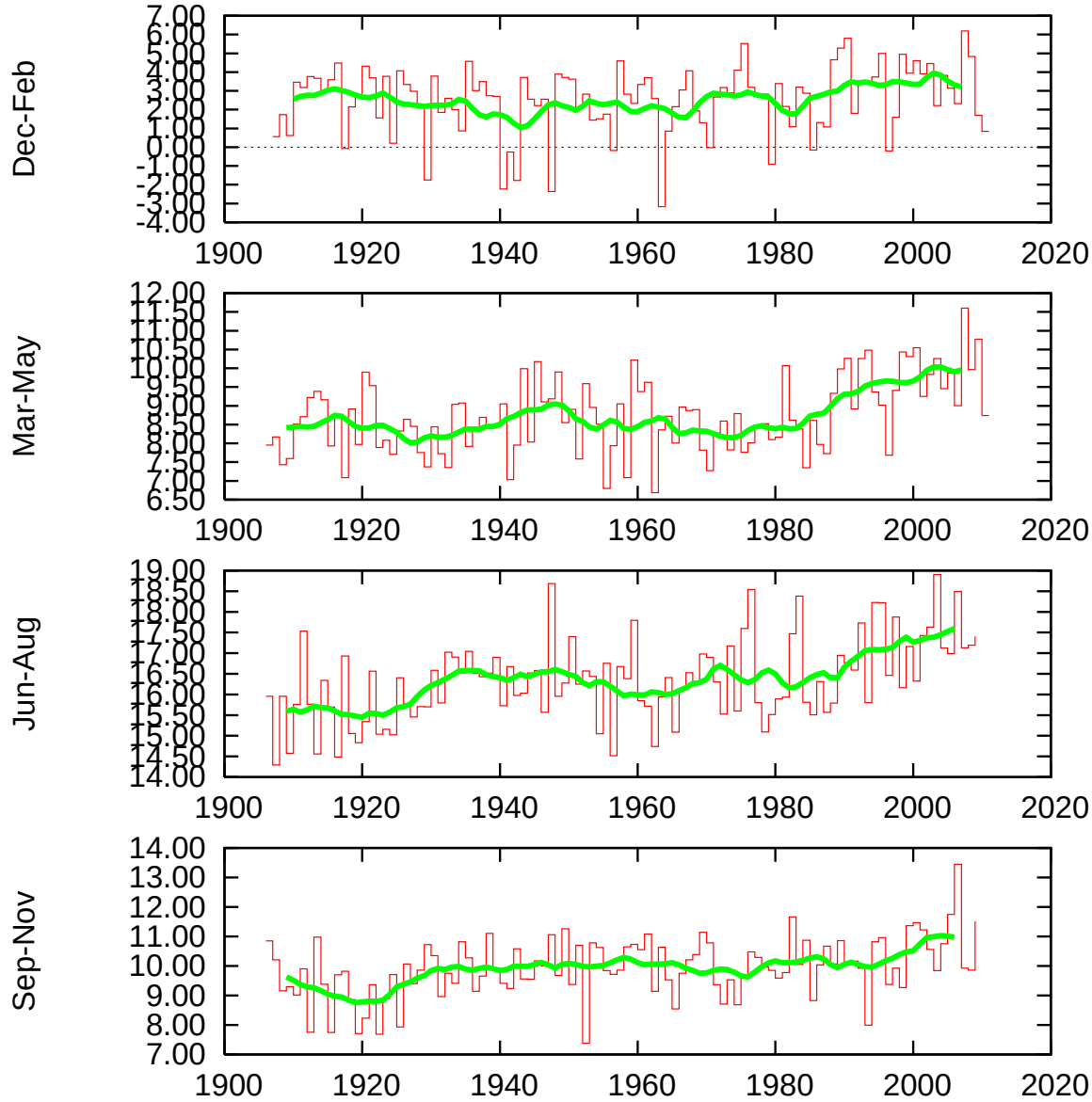
Oldenborgh, G.J. van, S.S. Drijfhout, A. van Ulden, R. Haarsma, A. Sterl, C. Severijns, W. Hazeleger and H. Dijkstra, *Western Europe is warming much faster than expected.* Climate of the Past, 2009, **5**, 1-12.

Oldenborgh, G.J. van, F. Doblas-Reyes, W. Hazeleger and B. Wouters, *Skill in the trend and internal variability in a multi-model decadal prediction ensemble*, in preparation

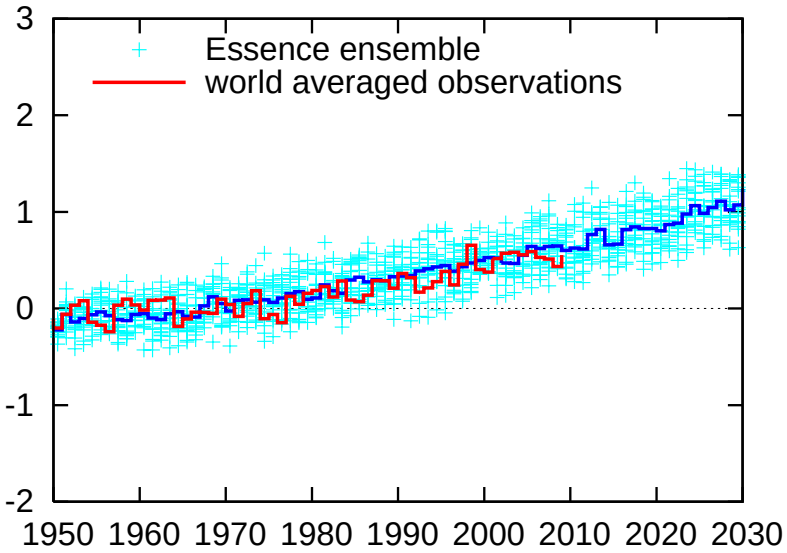
Temperatures at De Bilt 1706-2009



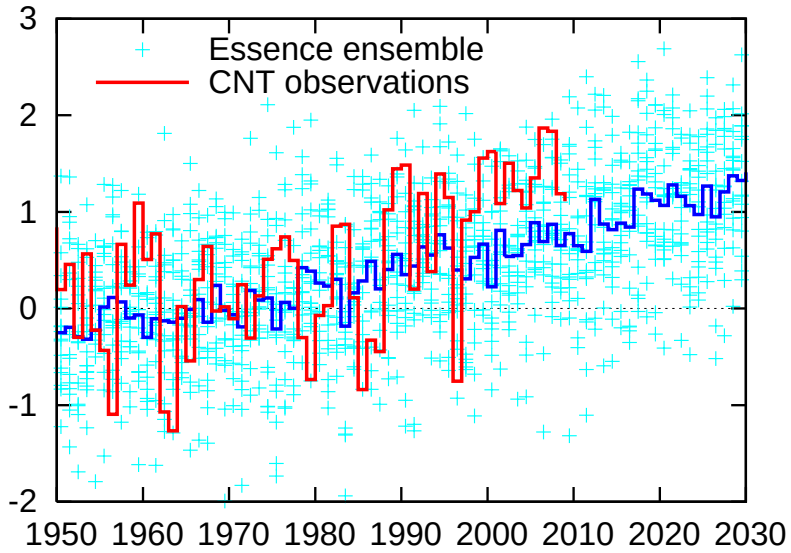
Central Netherlands Temperature 1906-2009



Comparison with climate model (17× ECHAM5)



Global mean temperature



Central Netherlands Temperature

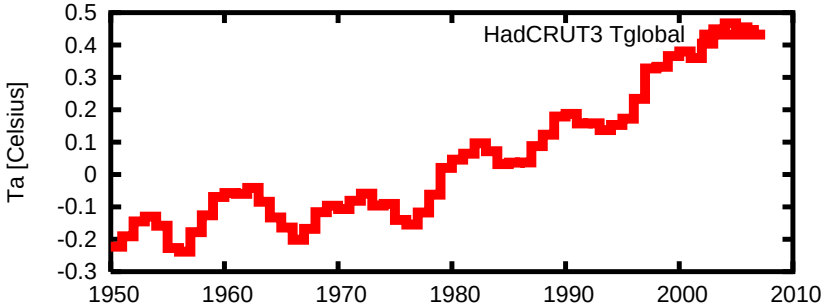
Definition of trend

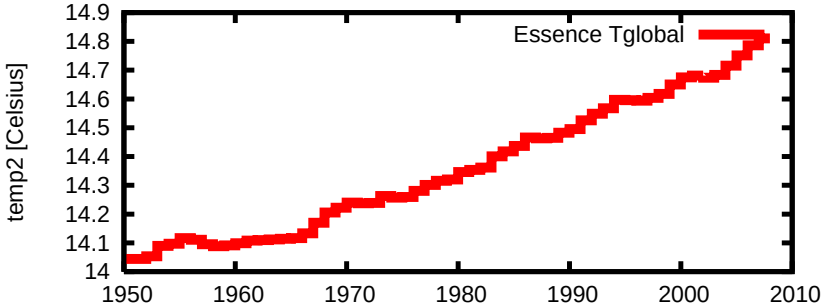
- The *difference of averaged periods* is very noisy due to the step functions at the edges of the intervals. 1997 shifting in and out of the averages makes a *big* difference.
- A *linear trend* is better, but still depends on the starting date. Starting in 1963 gives a different result from 1940 or 1975. Also, the trend *is* not linear.
- Therefore the trend is defined by the *regression against global mean temperature*. This method gives rise to smaller residuals.

$$T'(x, y, t) = A(x, y) T'^{(3)}_{\text{global}}(t) + \epsilon(x, y, t) \tag{1}$$

(A 3-yr running mean is applied to $T'_{\text{global}}(t)$ to filter out ENSO effects.)

Definition of trend





Fingerprint function.

Observations and models

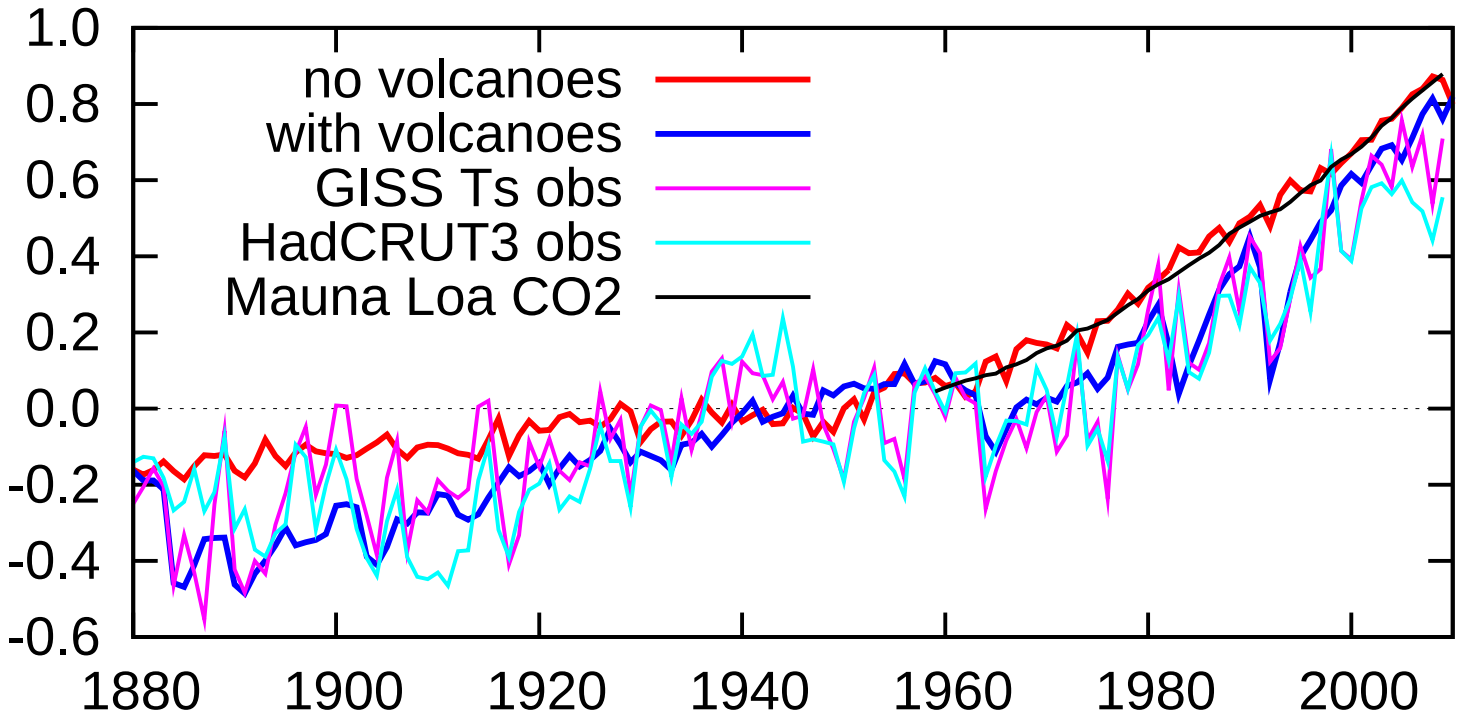
Observations:

- De Bilt homogenised temperature, Central Netherlands Temperature, CET
- CRUTEM3 and HadSST2, weighted according to fraction land/sea

GCM ensembles:

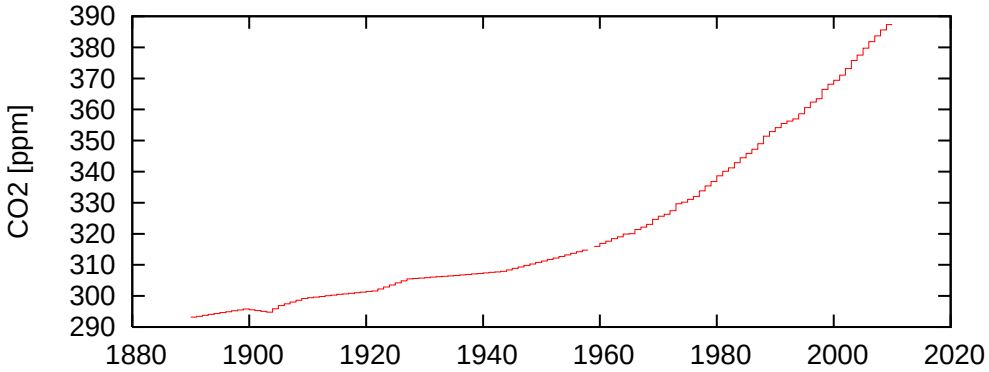
- ESSENCE: 17-member ECHAM5/MPI-OM 1950-2100 (20C3M, SRES A1B) T63,
- 5 models from CMIP3 with the most realistic circulation over Europe (ECHAM5/MPI-OM, GFDL CM2.1, MIROC 3.2 T106, CCCMA CGCM 3.1 T63, HadGEM1), see van Ulden and van Oldenborgh ACP 2006,
- full CMIP3 ensemble, 22 models (left out GISS-EH),
- UKMO QUMP perturbed physics ensemble, 17 members, 1850-2100, 20C3M/SRES A1B.

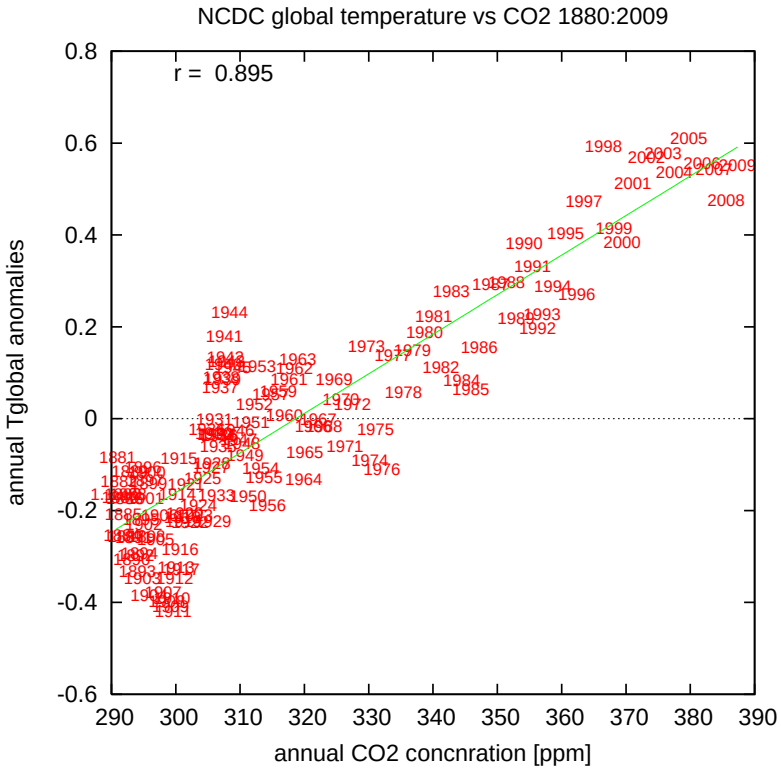
Global mean temperature



- Without volcanoes, T_{global} follows the CO_2 concentration
- With volcanoes the models track T_{global} quite well.

Forecasting global mean temperature

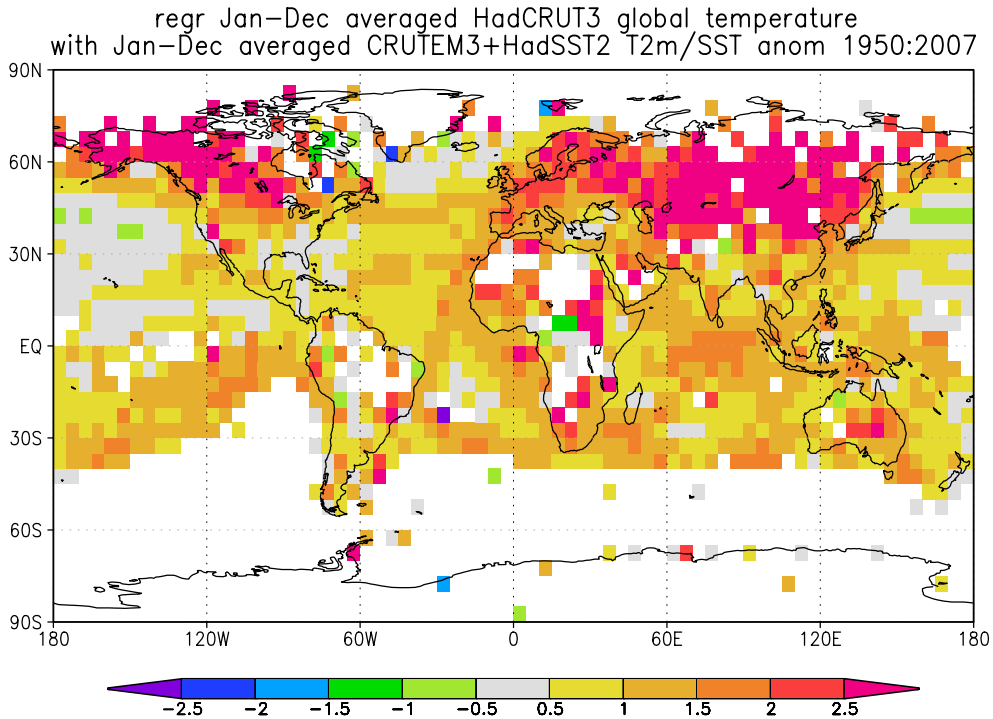




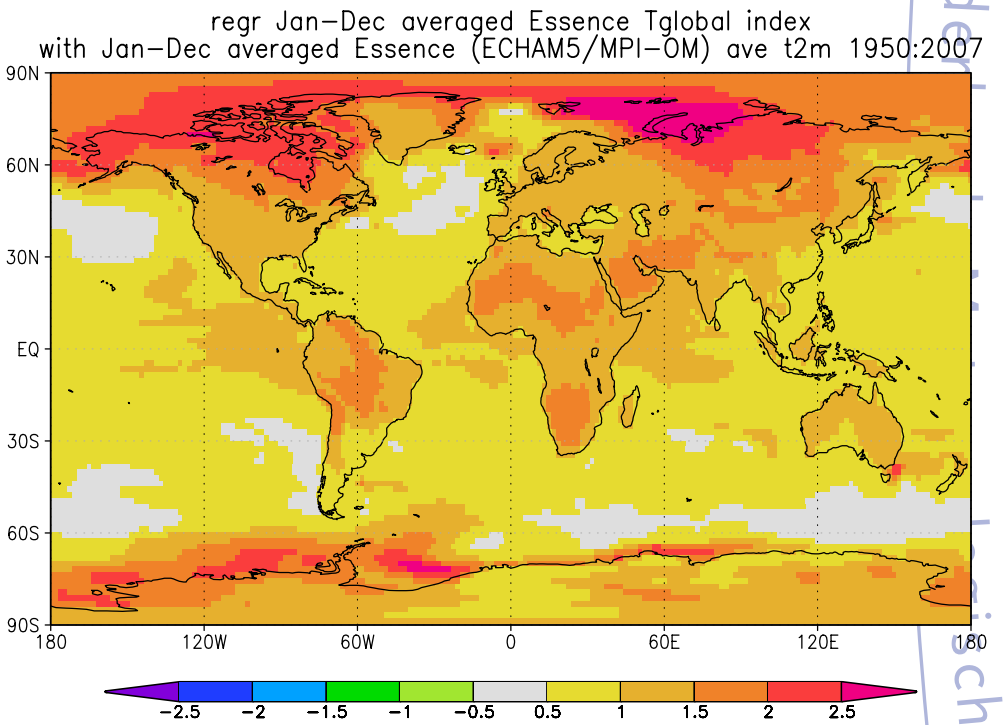
- CO_2 concentrations increase by 1.94(2) ppm/yr since the collapse of communism,
- Global mean temperature rises with 0.0086(4) K/ppm over 1880-2009,
- Hence we expect T_{global} to rise with 0.017(1) K/yr.

Local vs global warming

CRUTEM3+HadSST2 observations



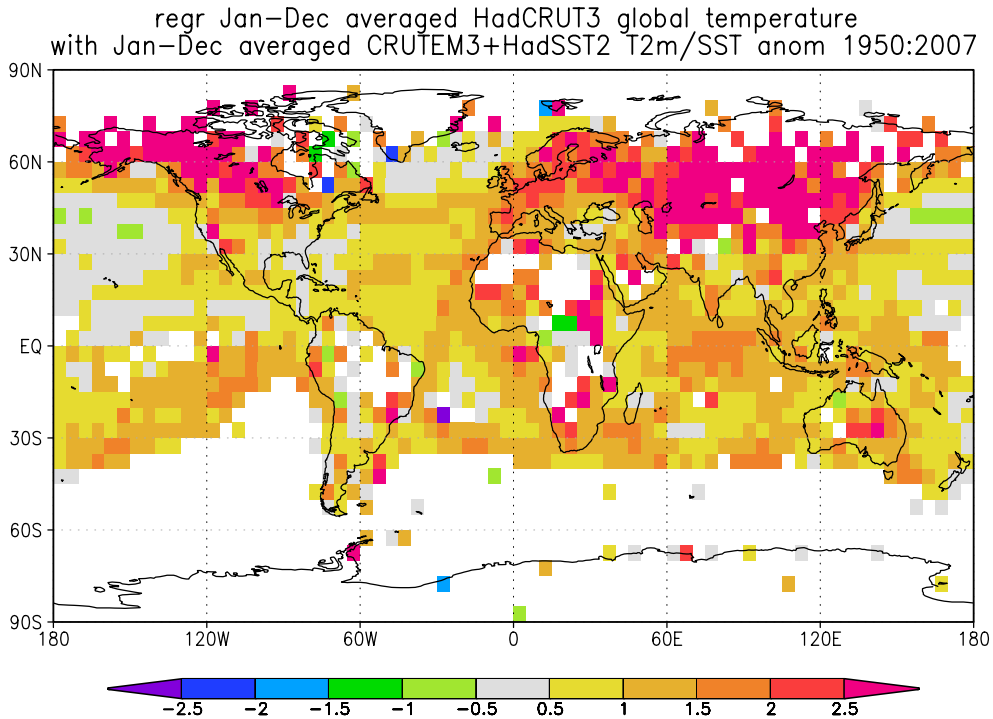
ESSENCE, 17 runs with ECHAM5/MPI-OM



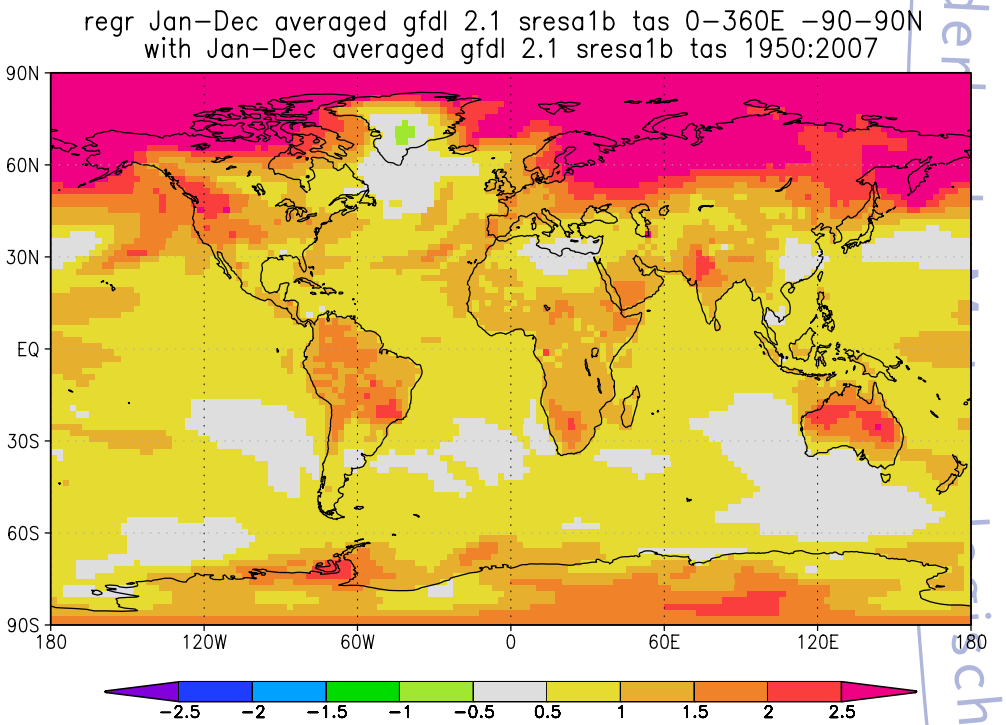
Main features are very similar, how about the details?

Local vs global warming

CRUTEM3+HadSST2 observations



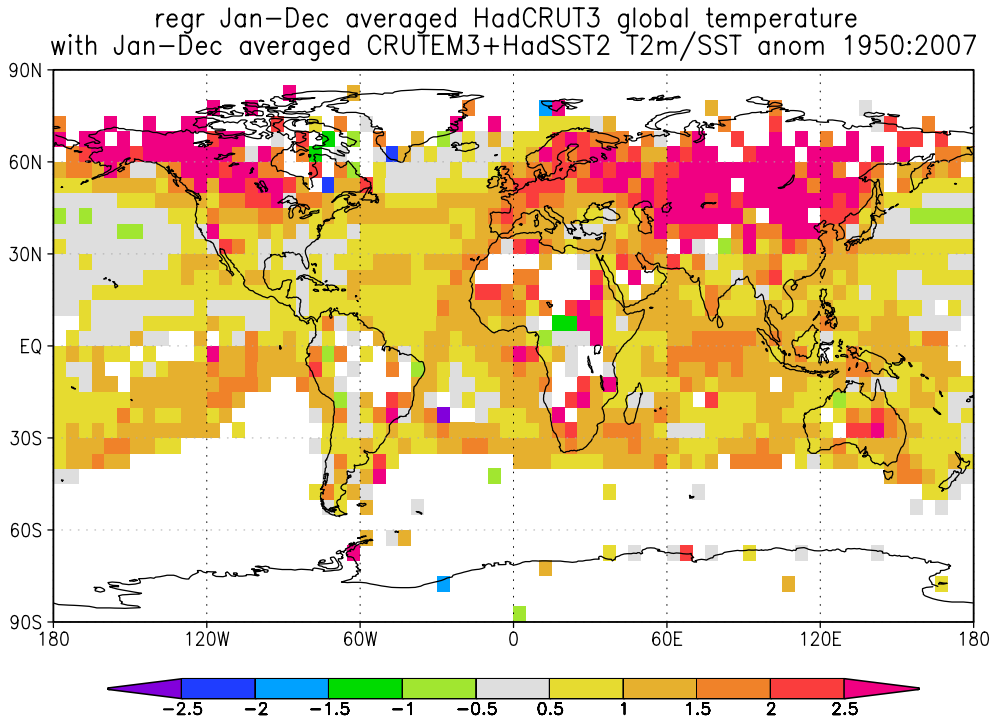
GFDL CM 2.1



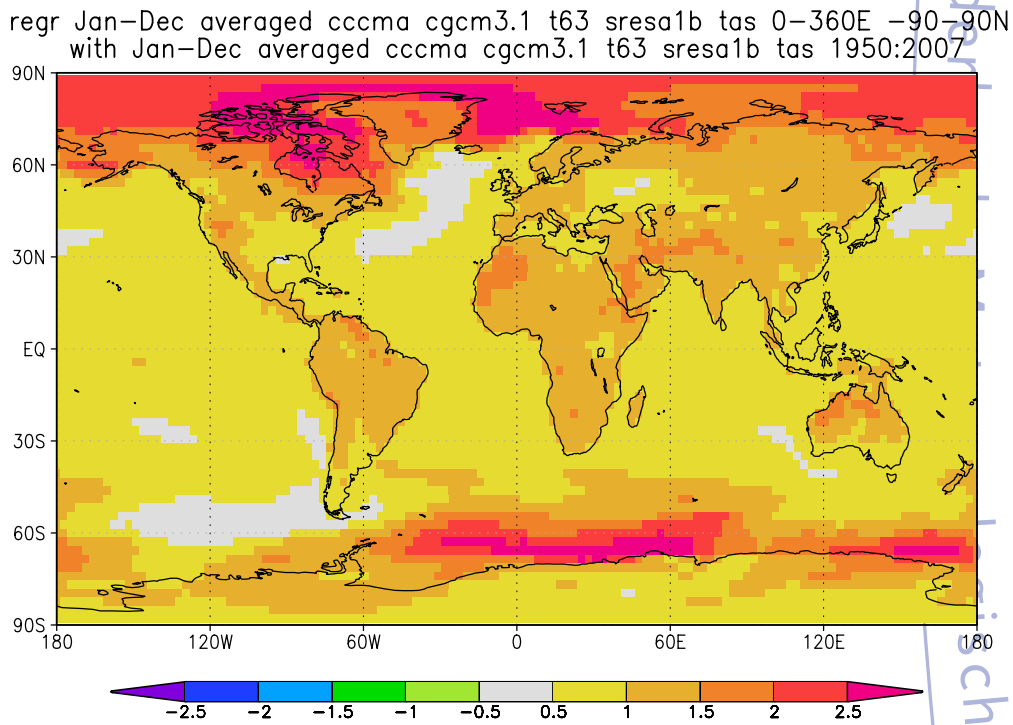
Main features are very similar, how about the details?

Local vs global warming

CRUTEM3+HadSST2 observations



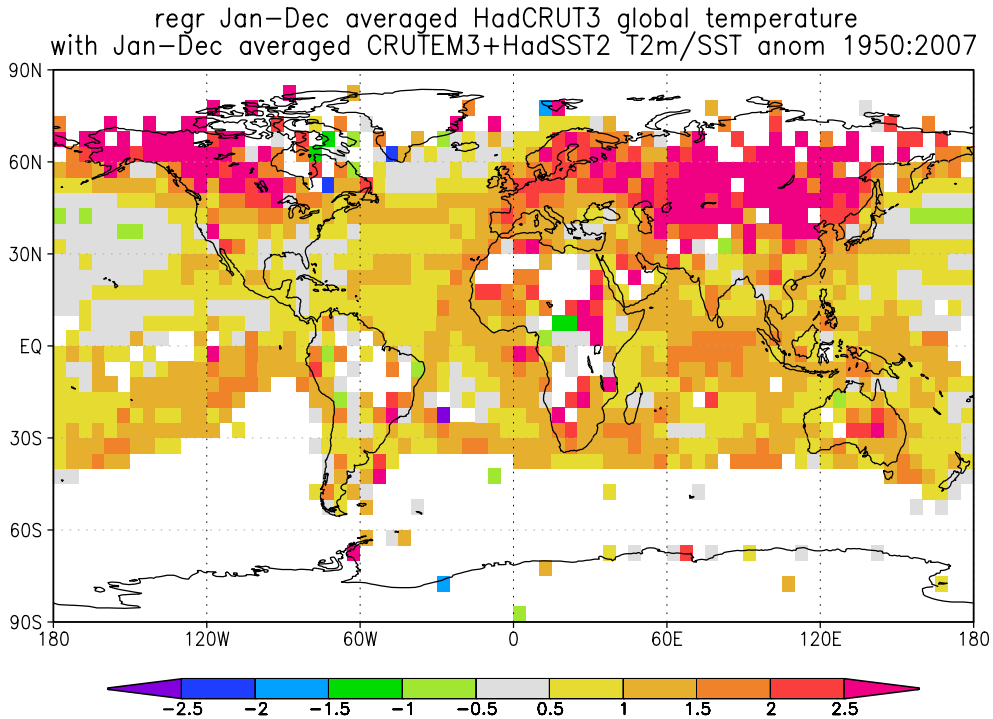
CCCMA CGCM 3.1 T63



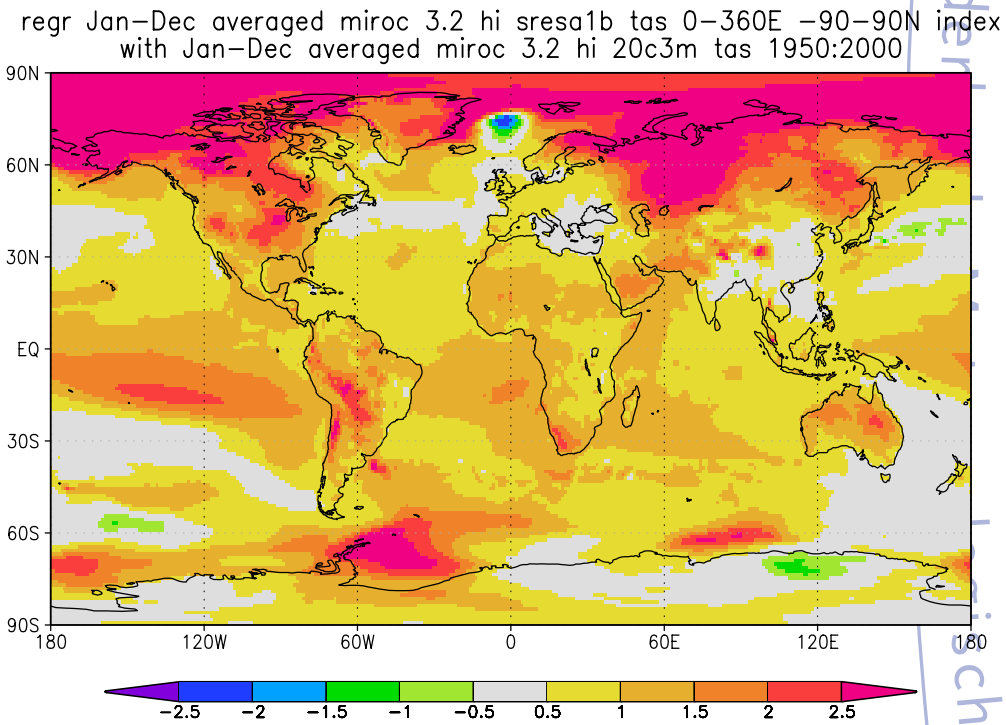
Main features are very similar, how about the details?

Local vs global warming

CRUTEM3+HadSST2 observations



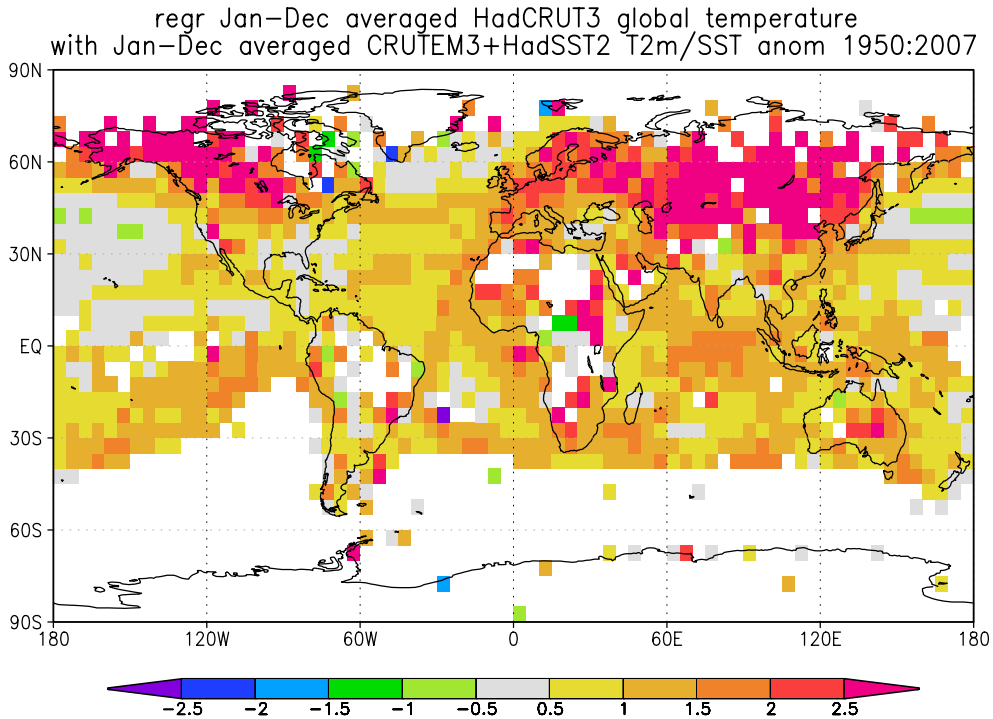
MIROC 3.2 T106



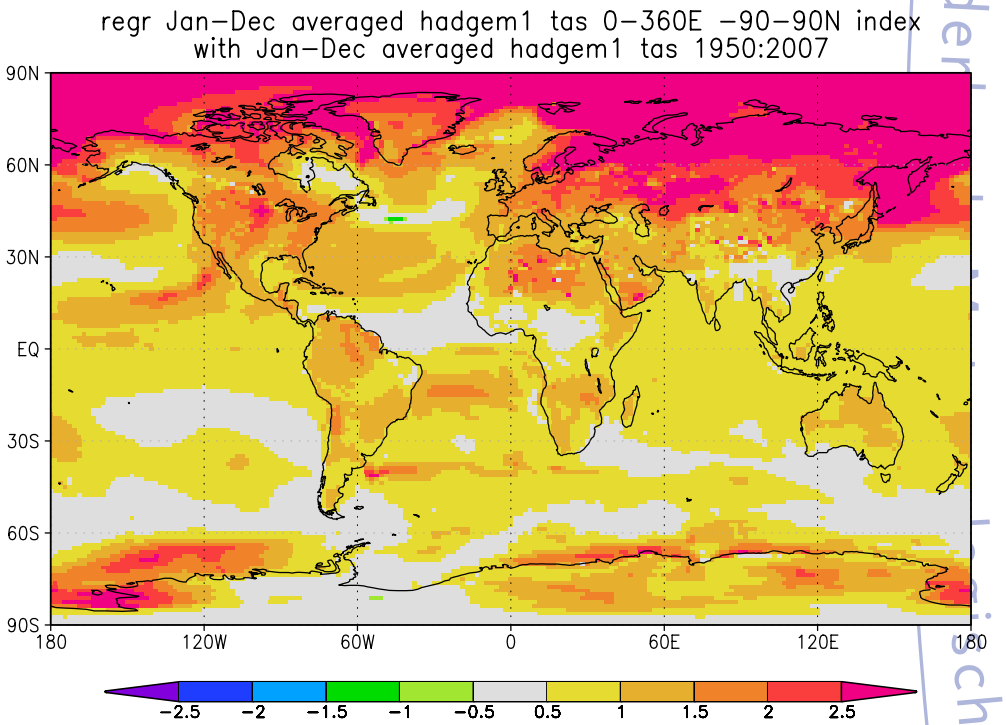
Main features are very similar, how about the details?

Local vs global warming

CRUTEM3+HadSST2 observations



UKMO HadGEM1



Main features are very similar, how about the details?

Significances of trend differences

There are two fundamentally different questions one can ask

Is the observed trend withn the PDF defined by the ensemble?

Assume that the models represent the mean and variability correctly, only use trend from observations.

Is the ensemble mean within the error bars of the observations?

Assume that the model mean is correct, and obtain the variability from the short observational record with a normality assumption.

Is the ensemble mean within the error bars of the observations?

Compute z-values from the regression estimates and their errors:

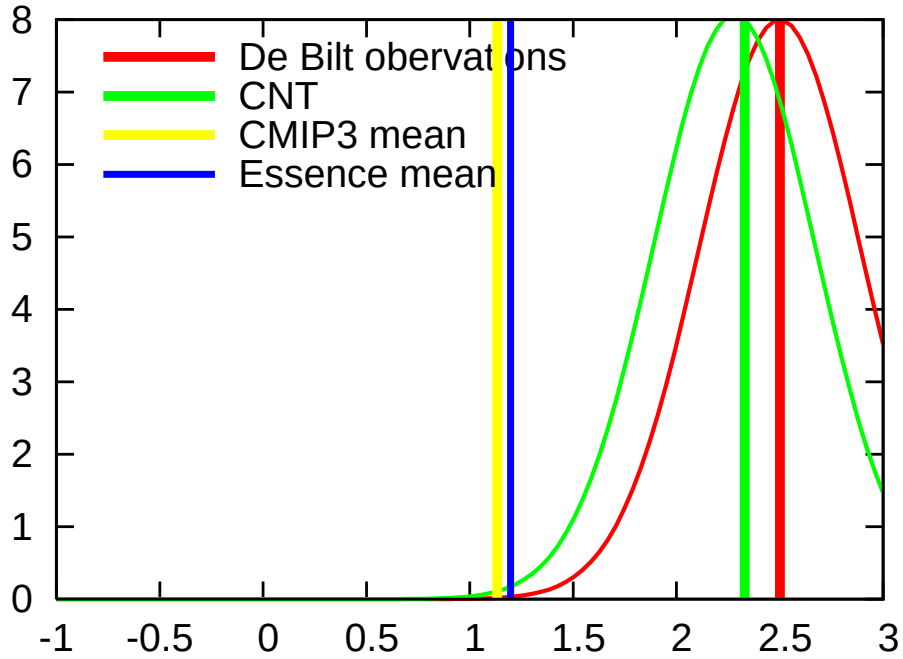
$$z = \frac{A_{\mathrm{obs}} - \overline{A}_{\mathrm{mod}}}{\sqrt{(\Delta A_{\mathrm{obs}})^2 + (\overline{\Delta A}_{\mathrm{mod}})^2/N}} \tag{2}$$

with N the number of ensemble members, the bar denotes the ensemble average. The standard errors ΔA are computed assuming a normal distribution of the trends A.

The normal approximation has been verified in the 17-member ECHAM5 ensemble, where the skewness of the 17 trend estimates is less than 0.2 in almost all areas where $z > 2$.

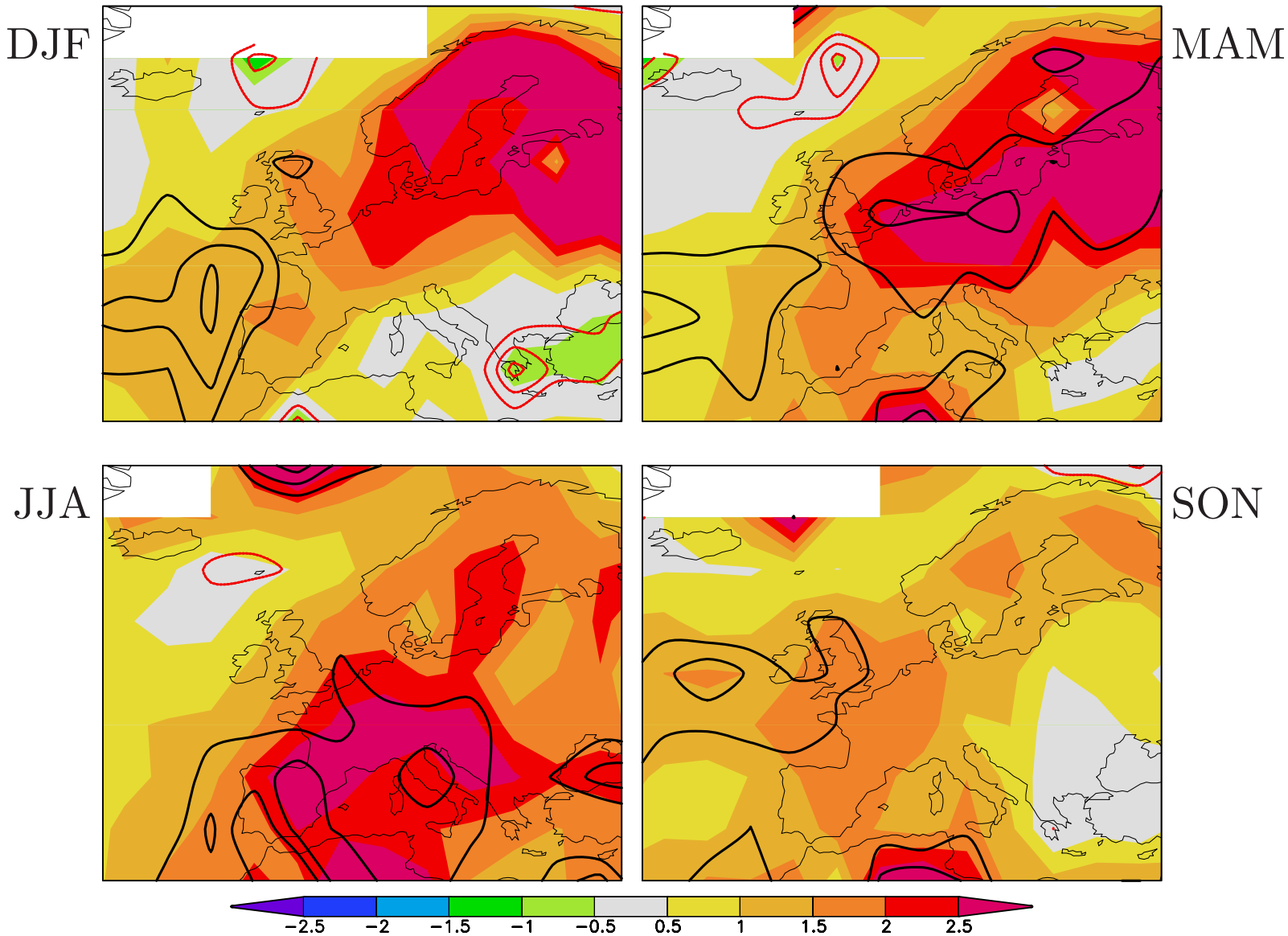
Serial correlations have been taken into account whenever significant.

Trends in the Netherlands, observations vs model



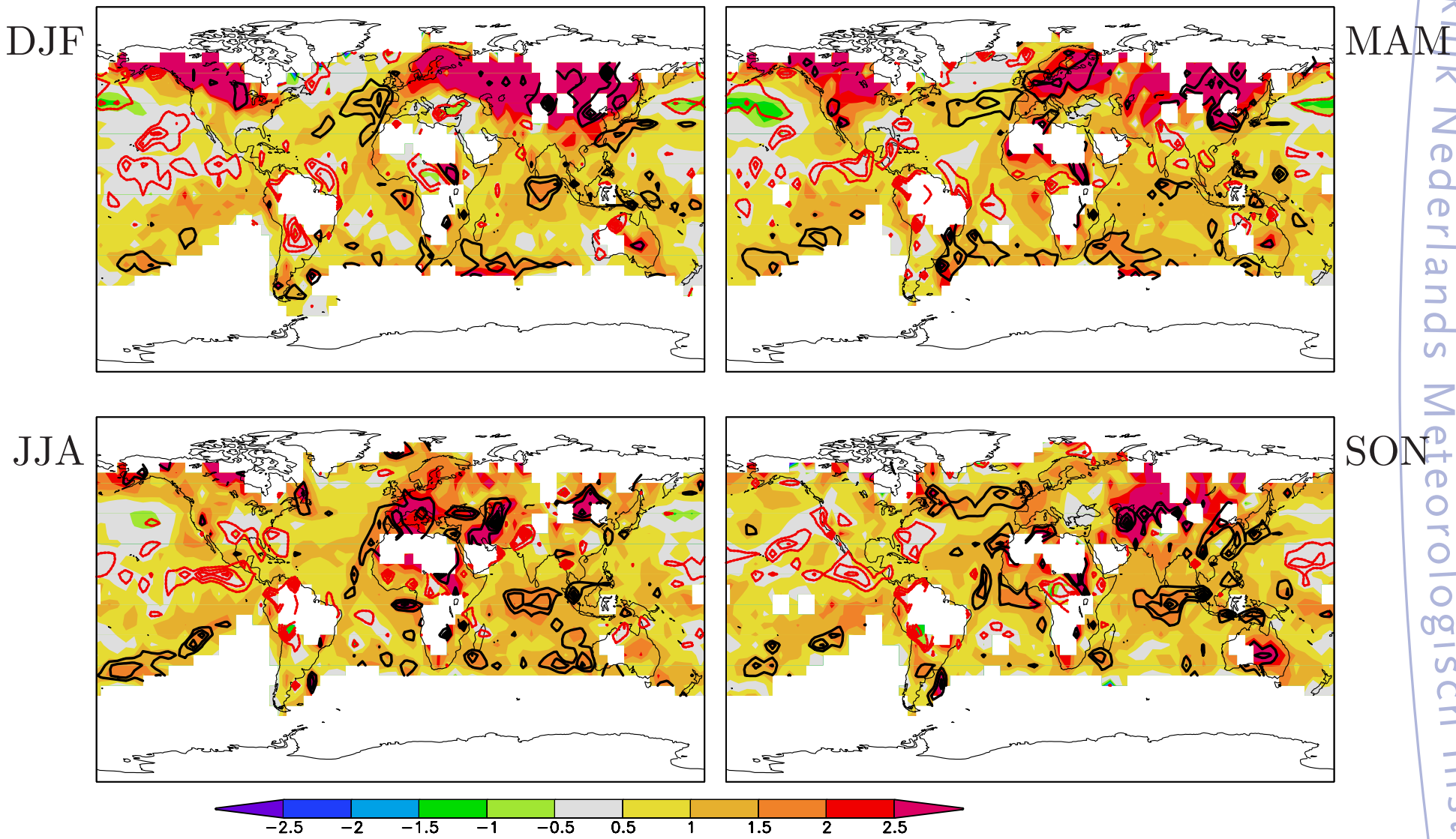
De Bilt (homogenised)	$A = 2.50 \pm 0.39$
Central Netherlands Temp.	$A = 2.33 \pm 0.38$
Essence	$A = 1.24 \pm 0.09 \Rightarrow z = 2.8 \quad (p \sim 0.012)$

- Is the ESSENCE mean trend within the error bars of the observations?



- Contours at $z = \pm 2, 3, 4$

Is the ESSENCE mean trend within the error bars of the observations?



Contours at $z = \pm 2, 3, 4$

Is the observed trend within the PDF of the ensemble?

Compute the fraction q of models that have lower trends the the observed one

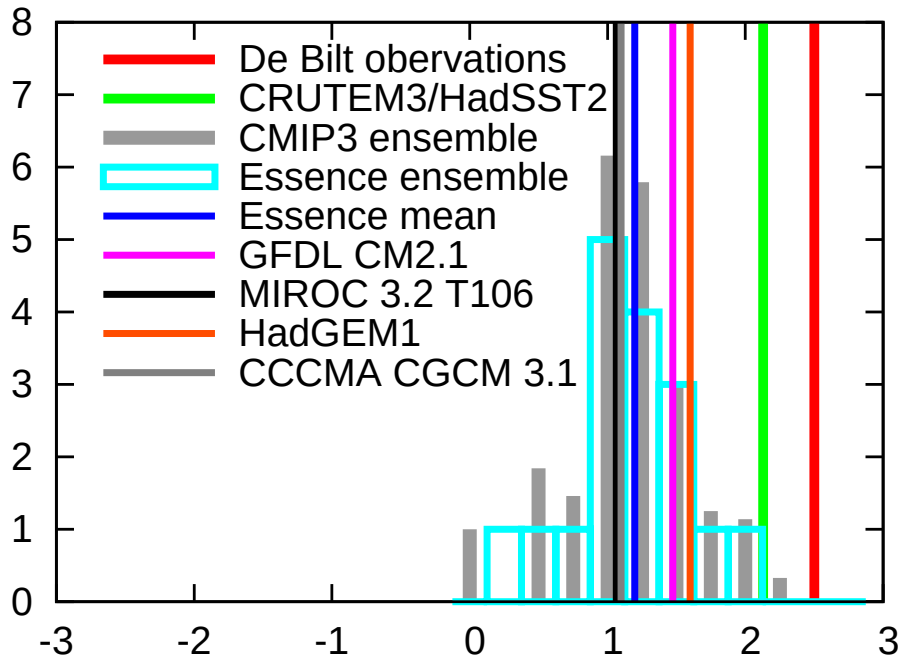
$$q = \frac{N + 1/2}{1 + N_{\mathrm{mod}}} \tag{3}$$

with N the number of models in the $N_{\mathrm{mod}} = 22$ model ensemble that have a trend lower than the observed one.

If there are $N_{\mathrm{run}} > 1$ runs for one model each run contributes $1/N_{\mathrm{run}}$ to N, so that the models are given equal weight.

Purple ($q > 0.975$) indicates that the observed trend is higher than all runs of all CMIP3 models simulate, in the red areas ($0.95 < q < 0.975$) one run of one model has a higher trend.

Trends in the Netherlands, observations vs model



De Bilt (homogenised) $A = 2.50 \pm 0.39$

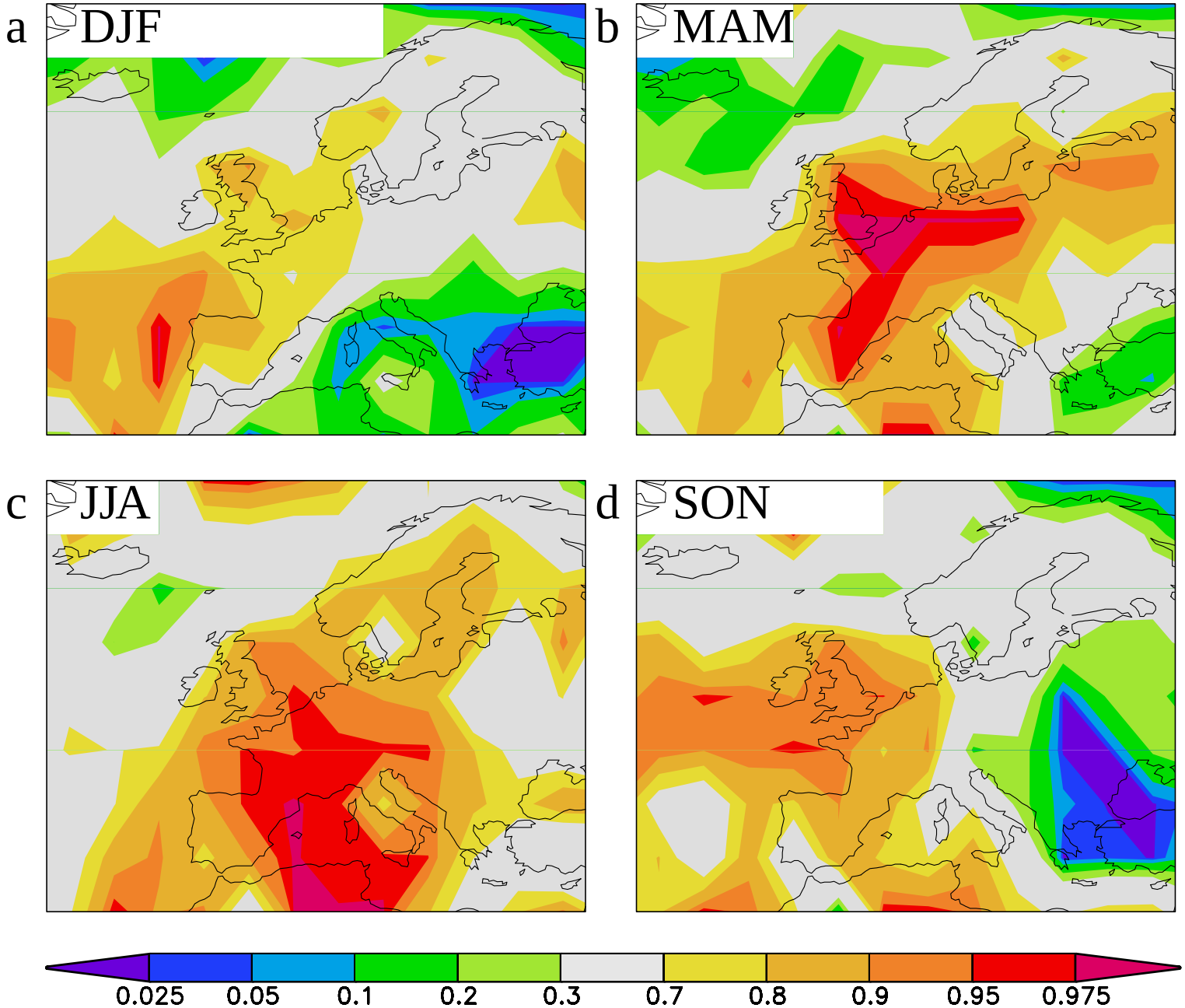
Central Netherlands Temp. $A = 2.23 \pm 0.39$

Above all 17 Essence ensemble members $(p \sim 0.028)$

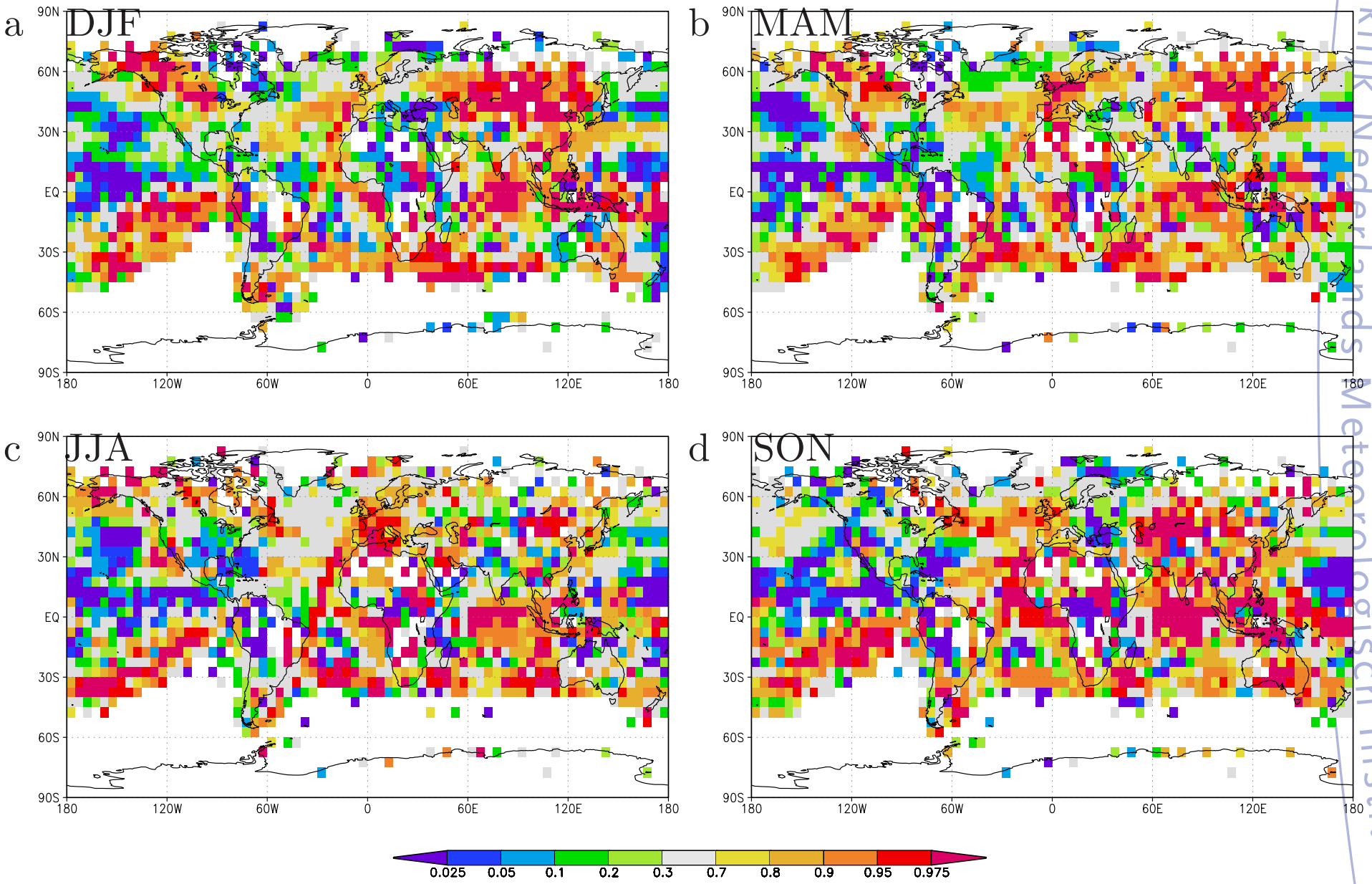
Above all 5 selected CMIP3 models

Above whole CMIP3 ensemble except one ensemble member $(p \sim 0.036)$

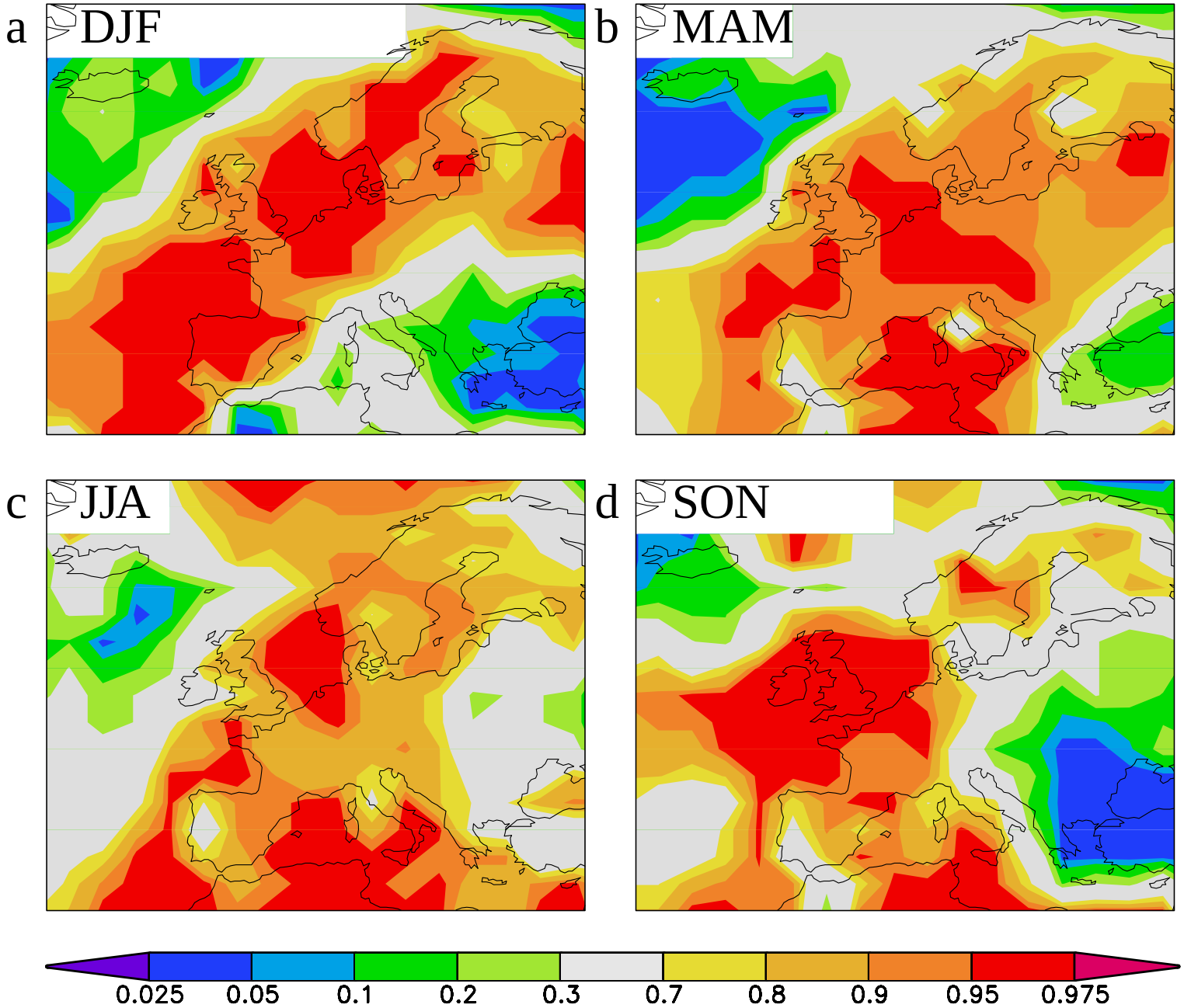
Is the observed trend within the PDF of the CMIP3 ensemble?



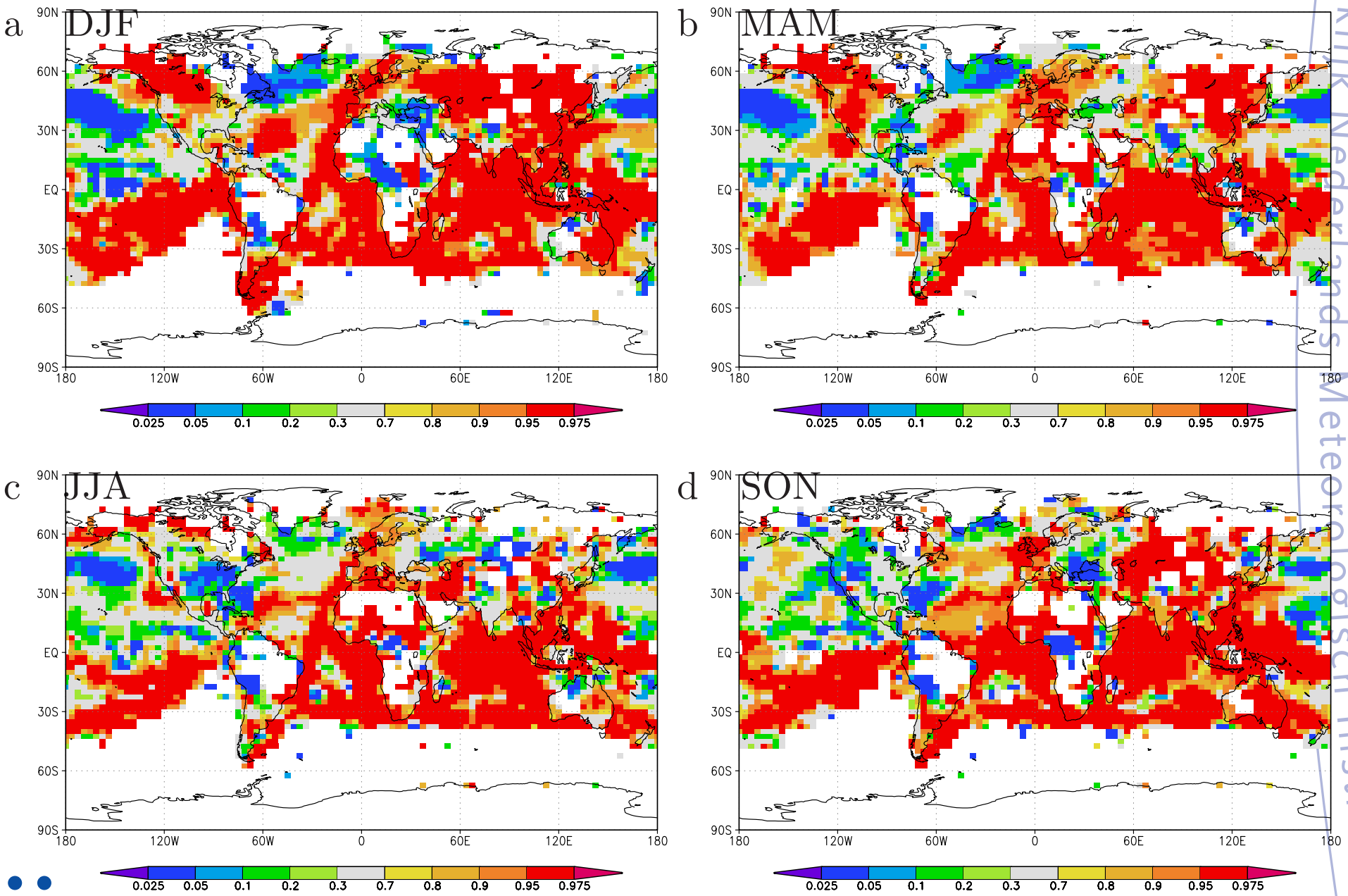
Is the observed trend within the PDF of the CMIP3 ensemble?



Is the observed trend within the PDF of a UKMO QUMP ensemble?



Is the observed trend within the PDF of a UKMO QUMP ensemble?



Higher resolution

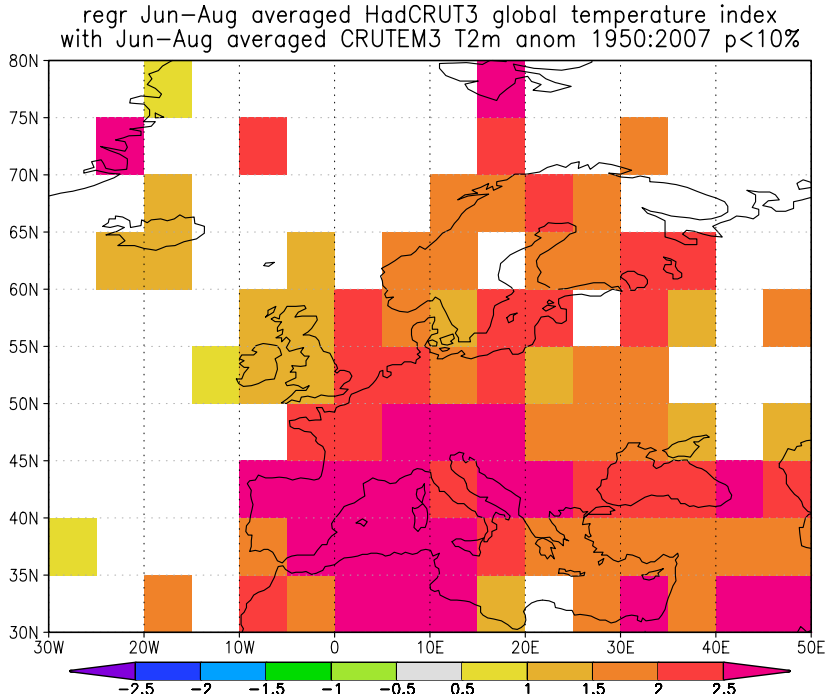
Observations:

- CRU TS 3
- ENSEMBLES 1.1

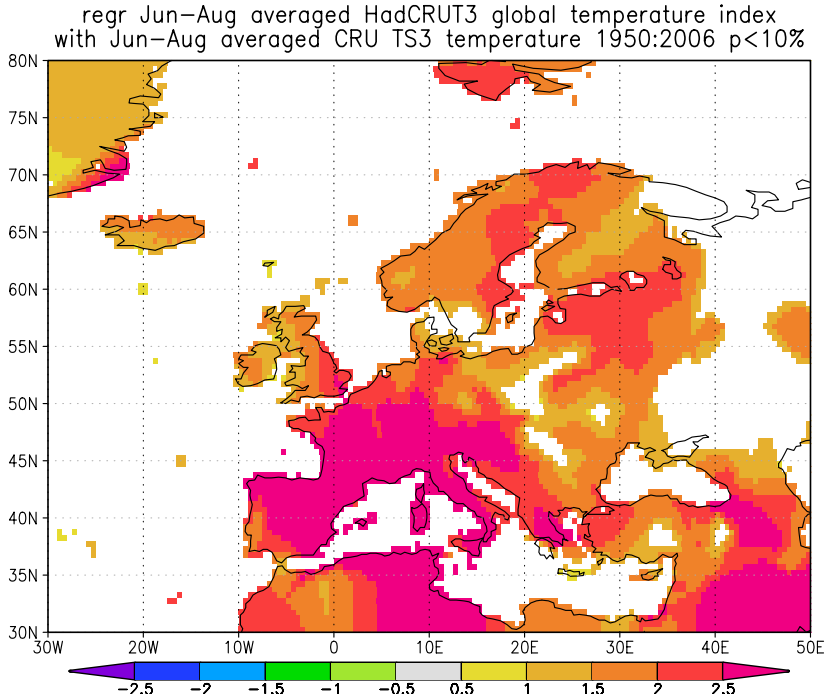
Models:

- ENSEMBLES RT2b: RCMs with GCM boundaries
- ENSEMBLES RT3: RCMs with ERA-40 boundaries

Observed trends do not depend greatly on the resolution

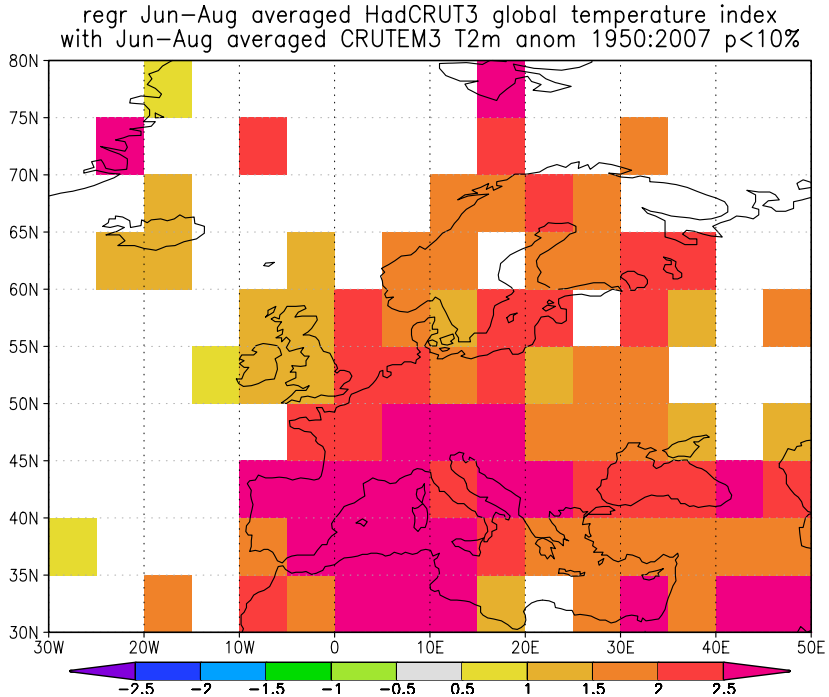


CRUTEM3 1950-2007

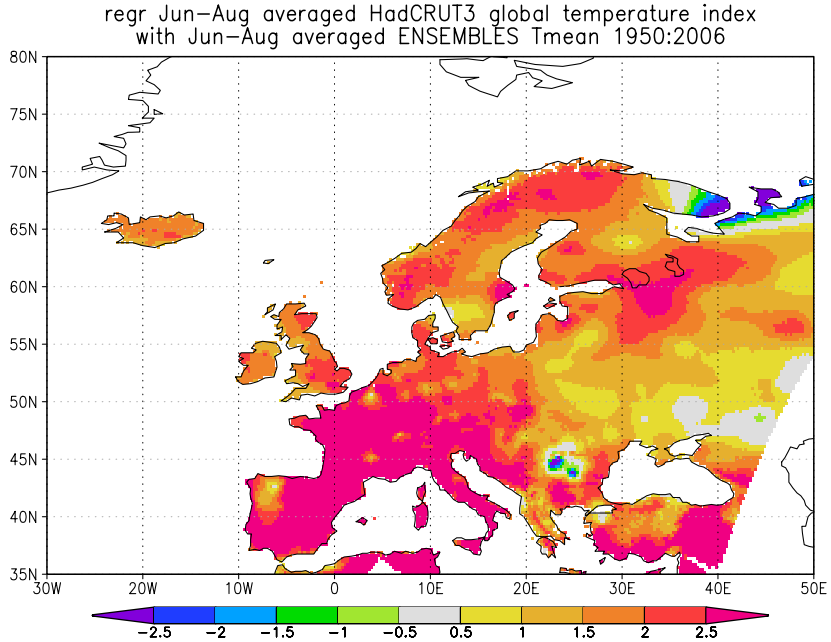


CRU TS 3 1950-2006

Observed trends do not depend greatly on the resolution

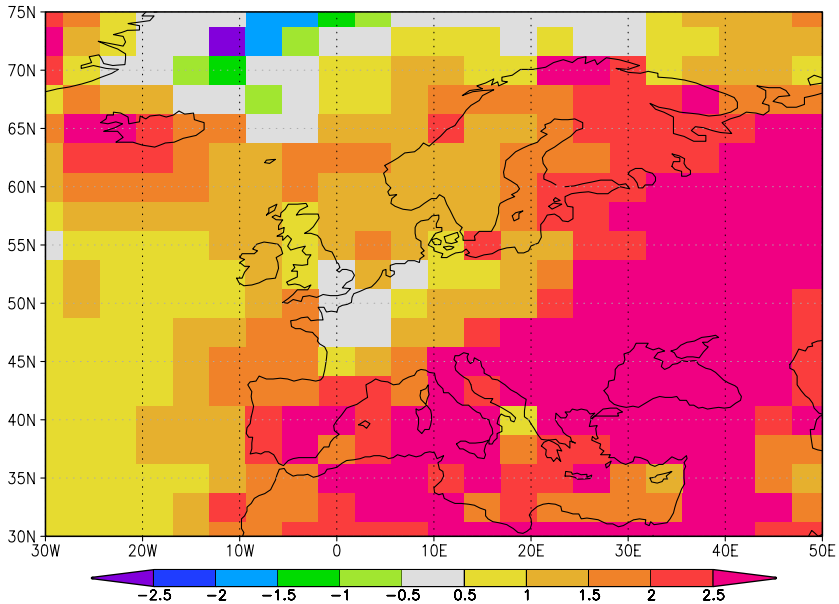


CRUTEM3 1950-2007

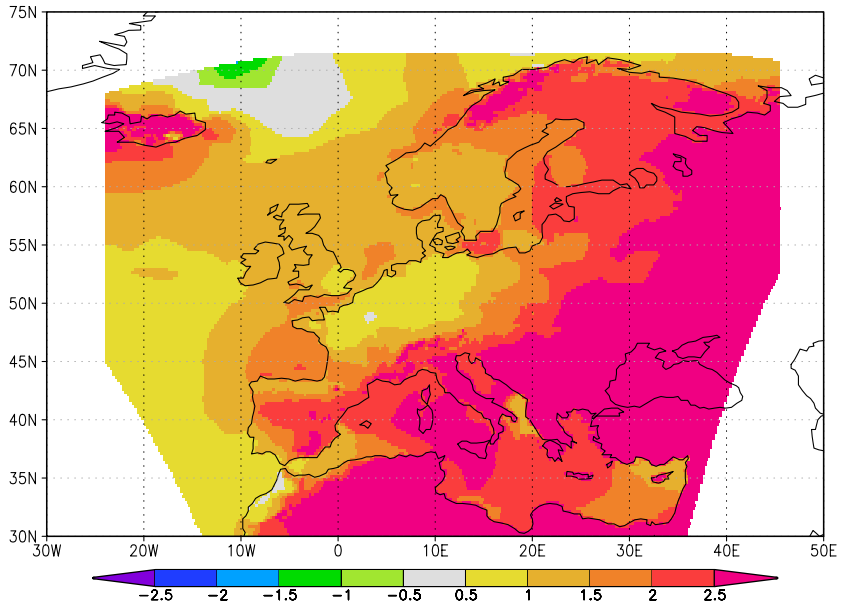


ENSEMBLES 1950-2006

Model trends do not depend greatly on the resolution

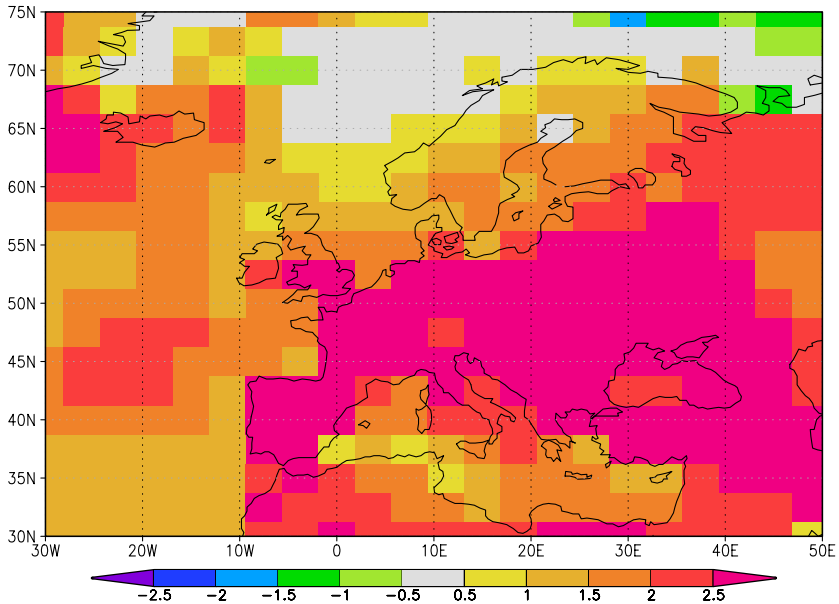


HadCM3 Q0

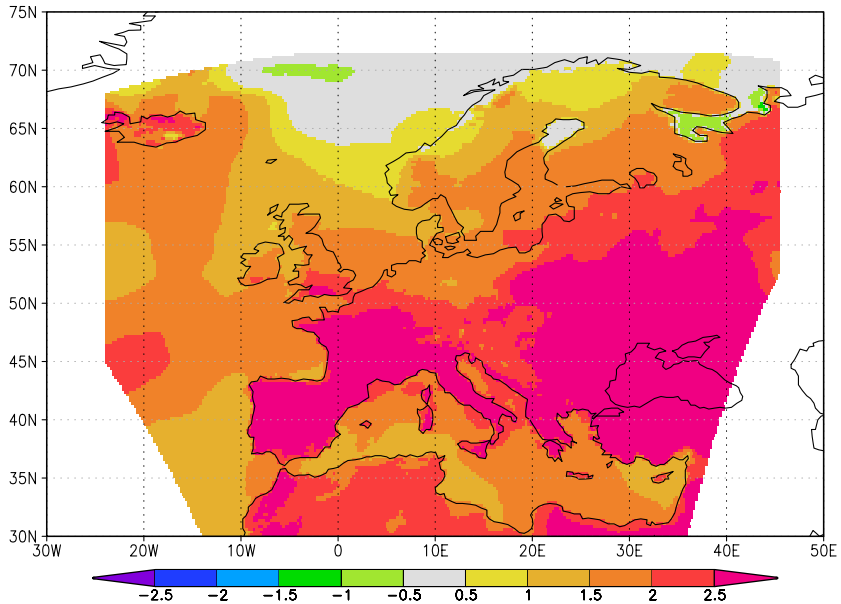


HadRM3 Q0

Model trends do not depend greatly on the resolution



HadCM3 Q16



HadRM3 Q16

Differences between observed and modelled temperature trends

- Do not depend strongly on resolution
- Highly significant in some seasons and regions
- Not due to weather or decadal climate fluctuations there

Physical causes of the differences in Europe

- Atmospheric circulation
- Ocean circulation
- Short-wave radiation (aerosols, clouds)
- Soil moisture
- Snow cover
- ...

Atmospheric circulation trends

Following van Oldenborgh & van Ulden (2003) and van Ulden & van Oldenborgh (2006) we construct a Very Simple Model (VSM) at each grid point

$$T'(t) = AT'_{\text{global}}(t) + \mathbf{B}\mathbf{u}'(t) + MT'(t-1) + \eta(t) \tag{4}$$

with $\mathbf{u}'(t) = (u_{\text{geo}}, v_{\text{geo}}, V_{\text{geo}})$ the geostrophic wind components: zonal, meridional and vorticity. M is a memory term, and the remainder is classified as noise $\eta(t)$, time t in months.

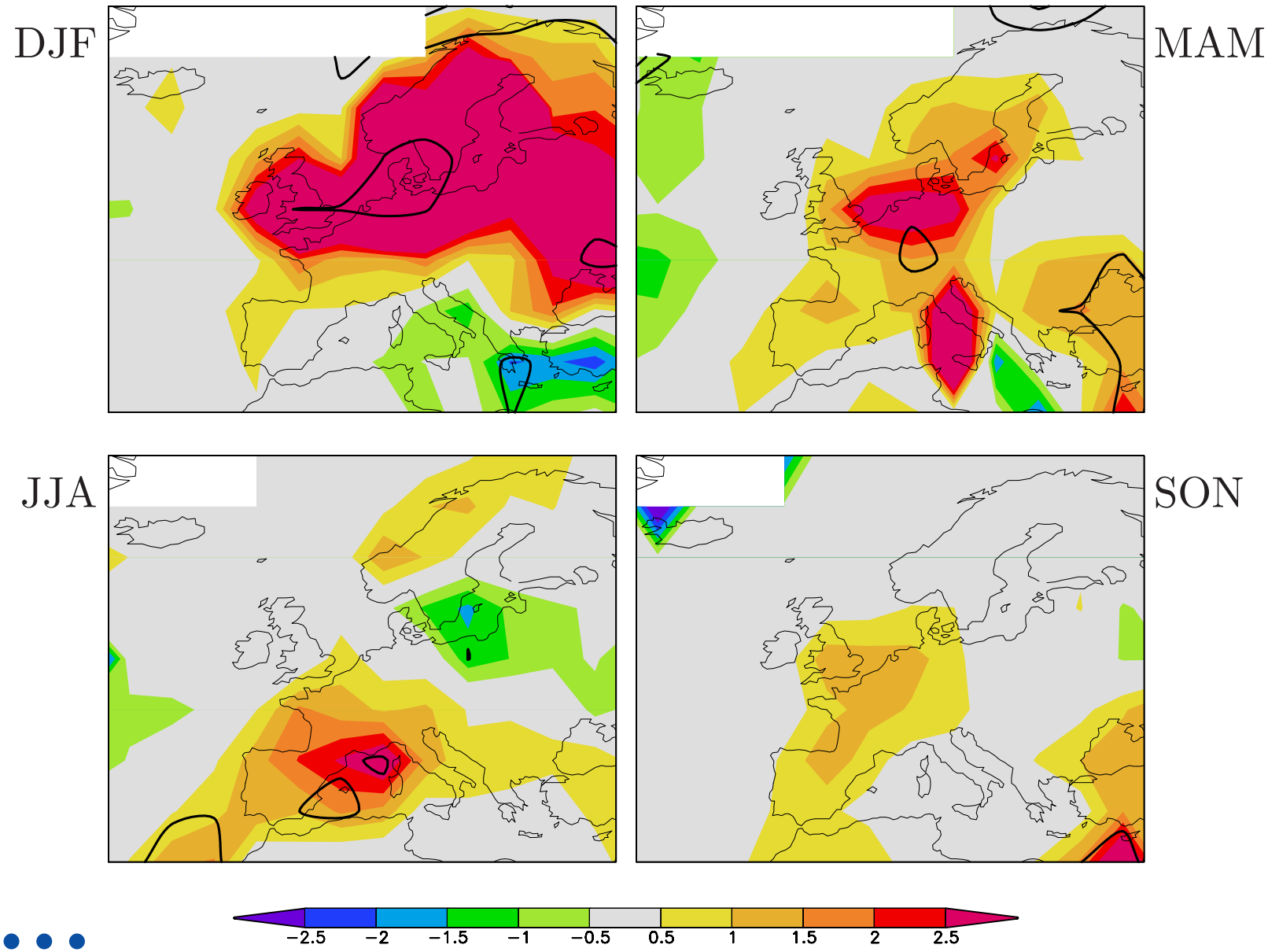
Define circulation (in)dependent temperature anomalies:

$$T'_{\text{circ}}(t) = \mathbf{B}\mathbf{u}'(t) \tag{5}$$

$$T'_{\text{ind}}(t) = AT'_{\text{global}}(t) + \eta(t) \tag{6}$$

Geostrophic winds and vorticity are computed from NCEP/NCAR reanalysis SLP.

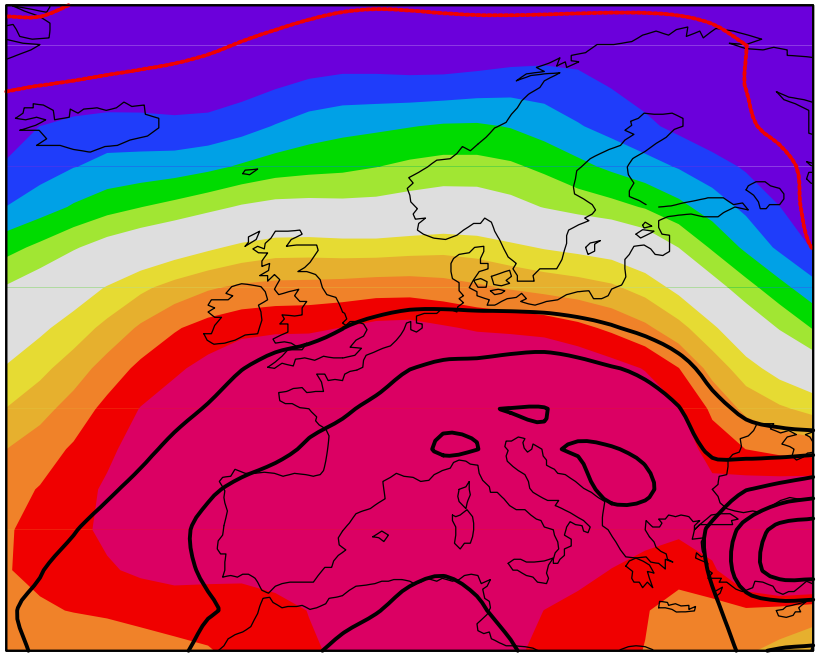
Trends in T_{circ}



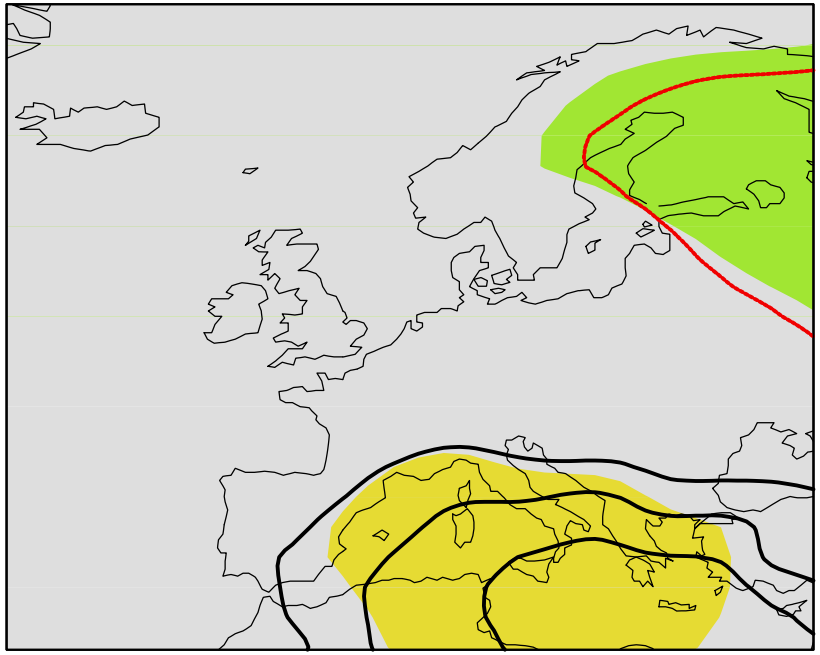
Trends in circulation

- Over most of Europe, the trends in the circulation-independent parts are largest in spring and summer; in winter the trends in the circulation dominate.

DJF



NCEP/NCAR SLP trend

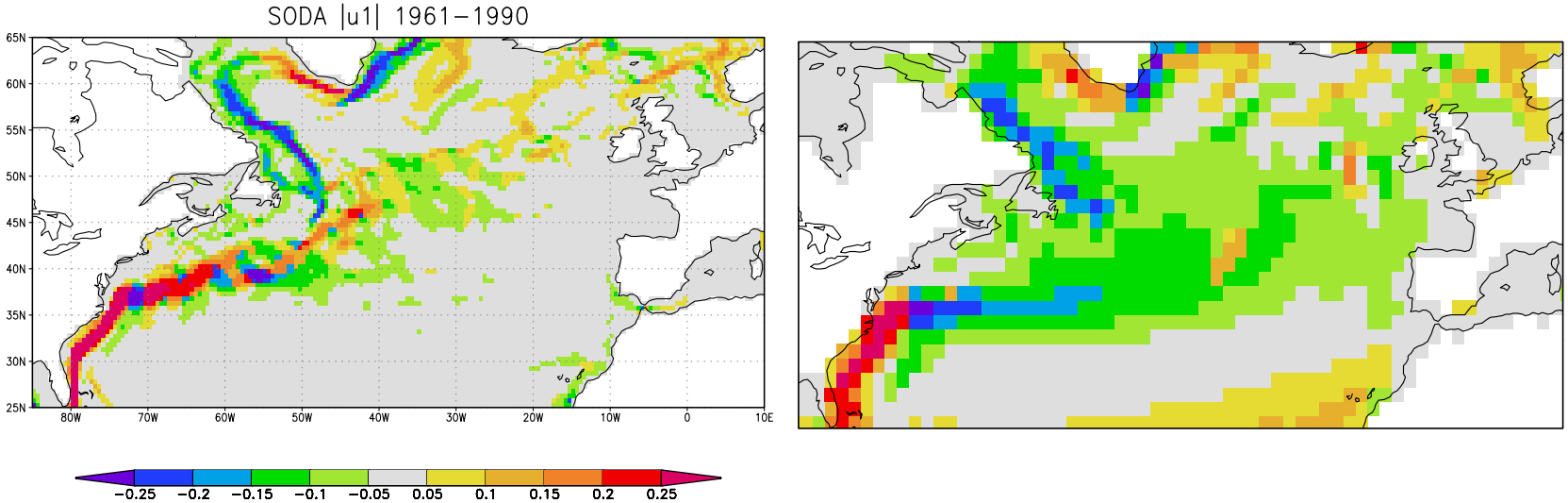


ECHAM5/MPI-OM SLP trend

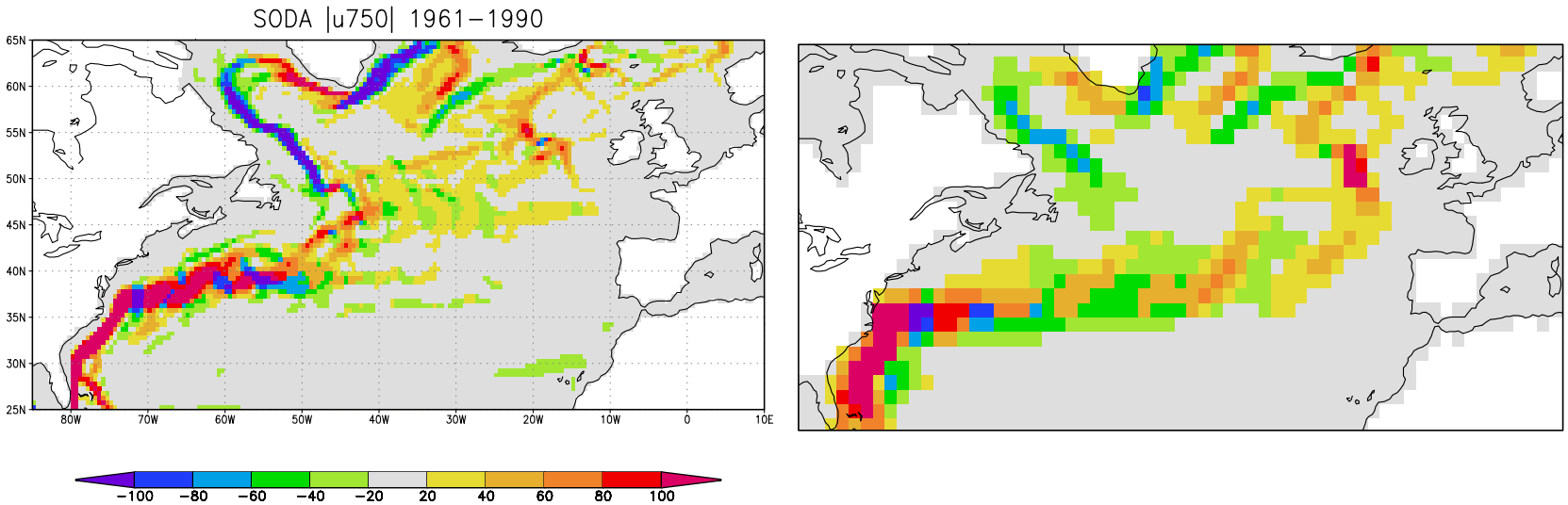
DJF

North Atlantic Ocean circulation bias

$u(10\mathrm{m})$



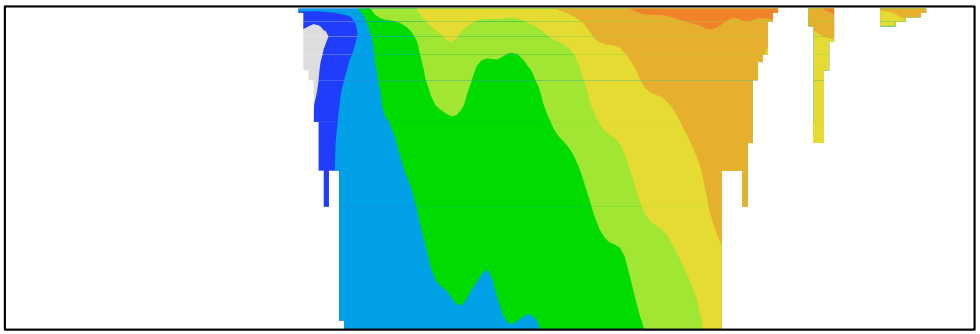
$\int_{z=0}^{750\mathrm{m}} u\, dz$



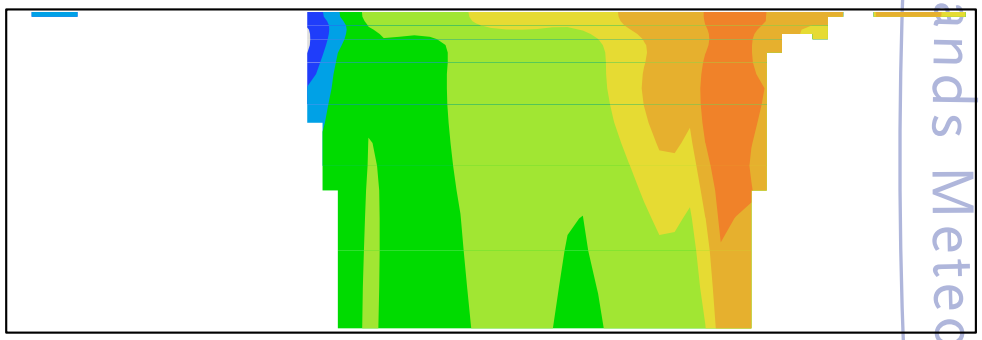
SODA 1.4.1

ECHAM5/MPI-OM

Section at 52°N



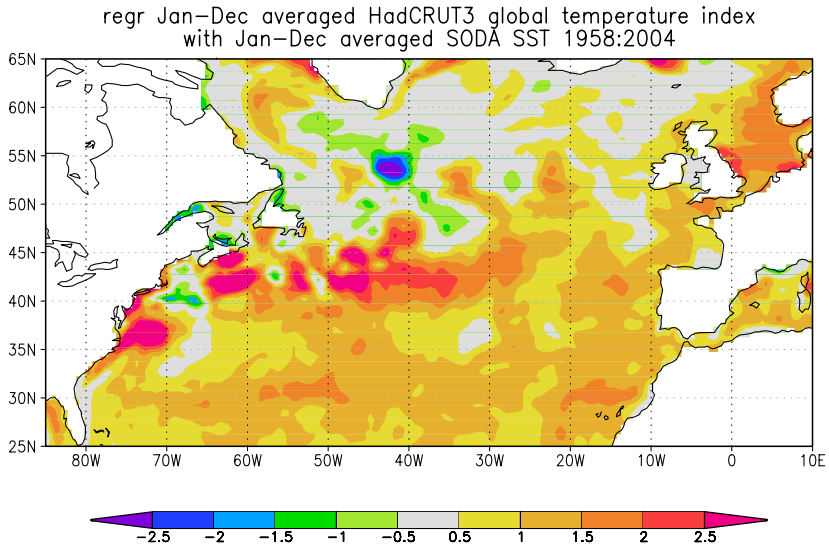
SODA 1.4.1



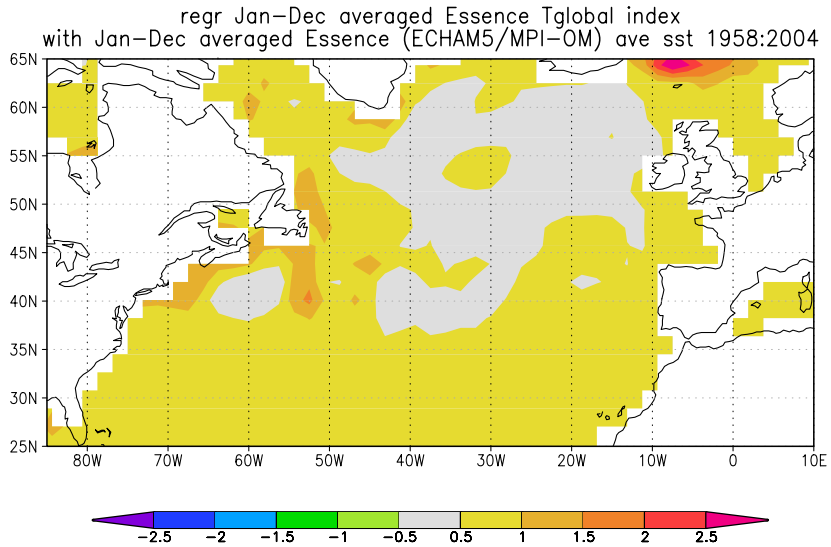
ECHAM5/MPI-OM

SST warming trends

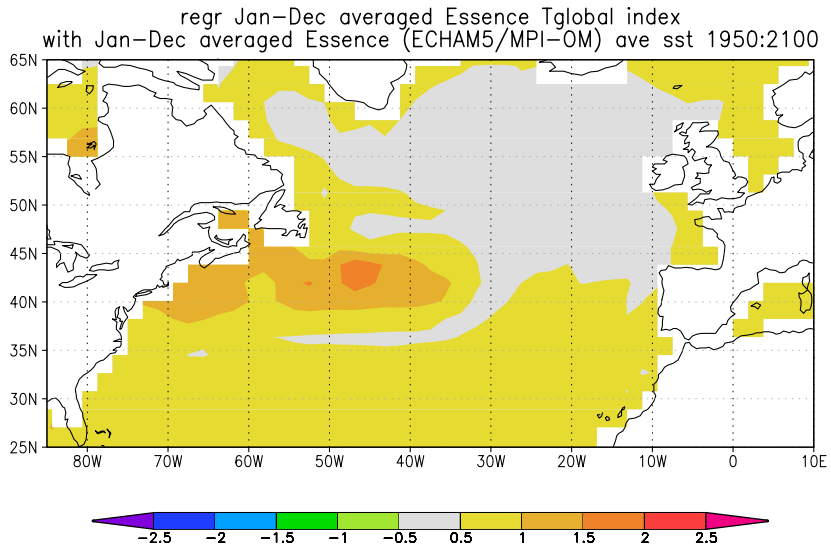
SODA-POP 1958-2004



ESSENCE 1958-2004

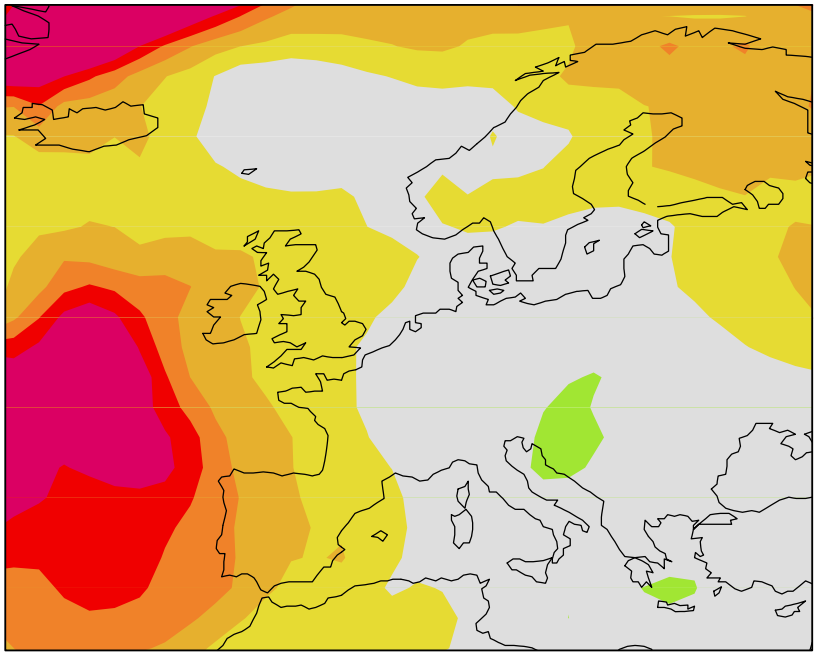


ESSENCE 1950-2100

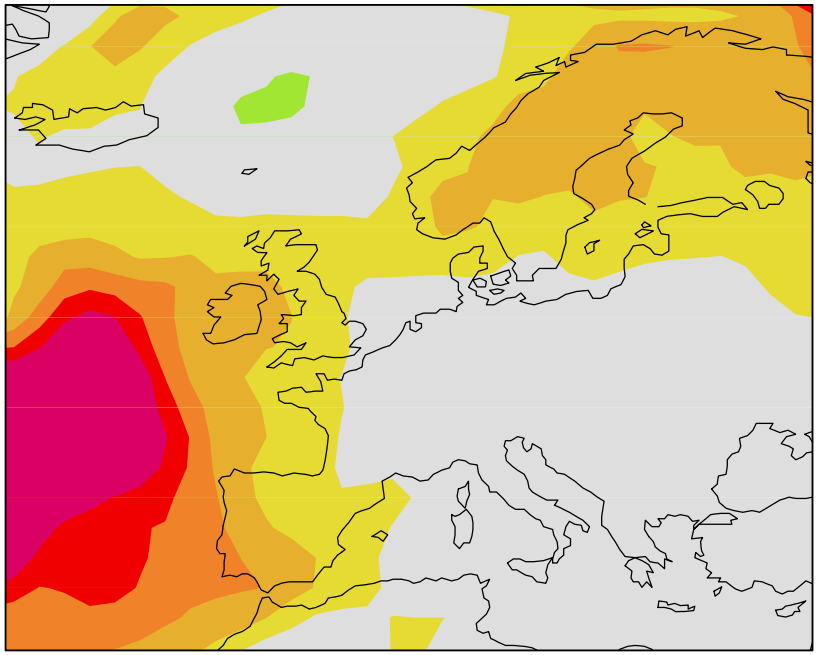


Modelled propagation to land

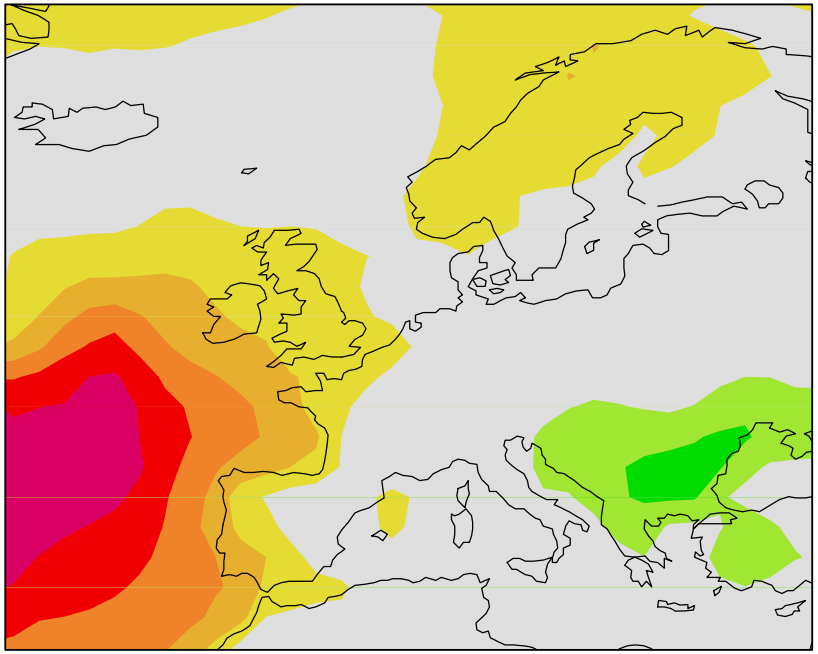
DJF



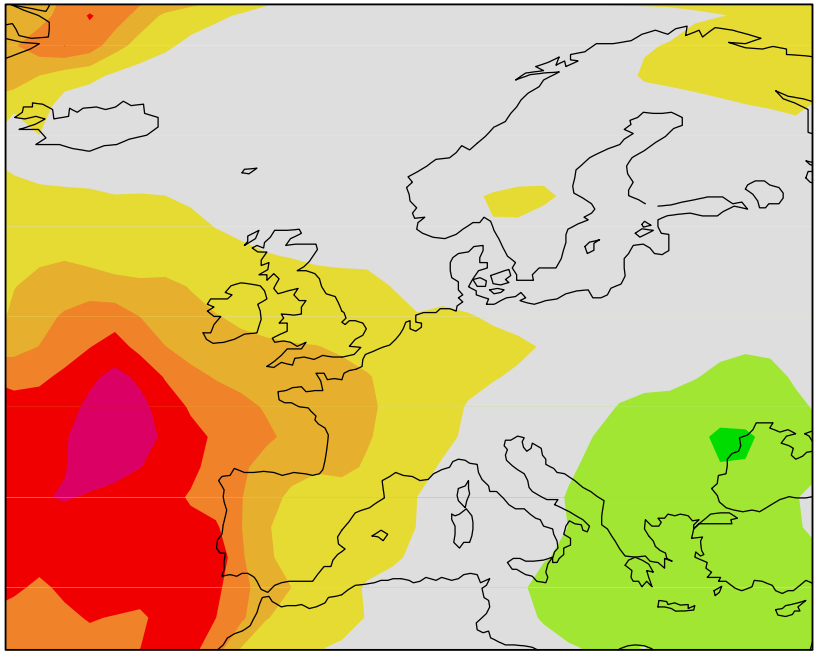
MAM



JJA



SON



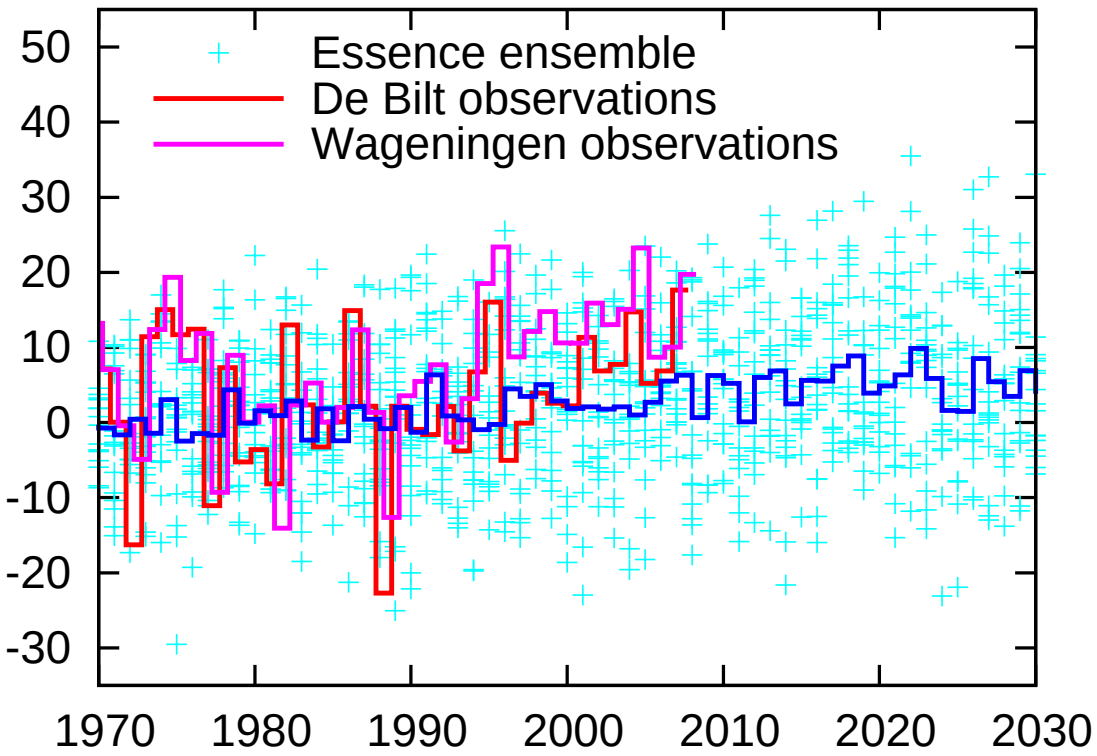
North Atlantic Ocean circulation bias

- The wrong flow of the North Atlantic ocean currents causes too deep mixed layers in the East Atlantic.
- SST trends are much smaller than observed in this area, also in other models.
- The bias in SST trends extends to the Atlantic coasts of Europe.

Note that the ensemble MOC decorrelates within 10 years, so all model decadal variations are sampled.

Short-wave radiation in spring and summer

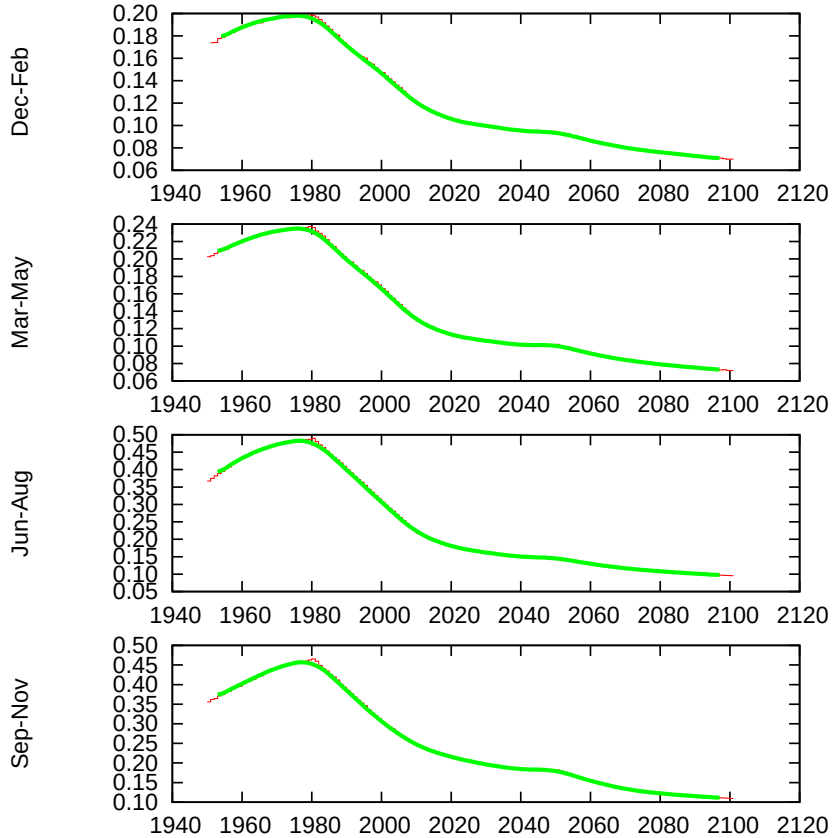
Global short-wave observations in the Netherlands are reasonably reliable from the early 1970s onwards. Corrected for circulation.

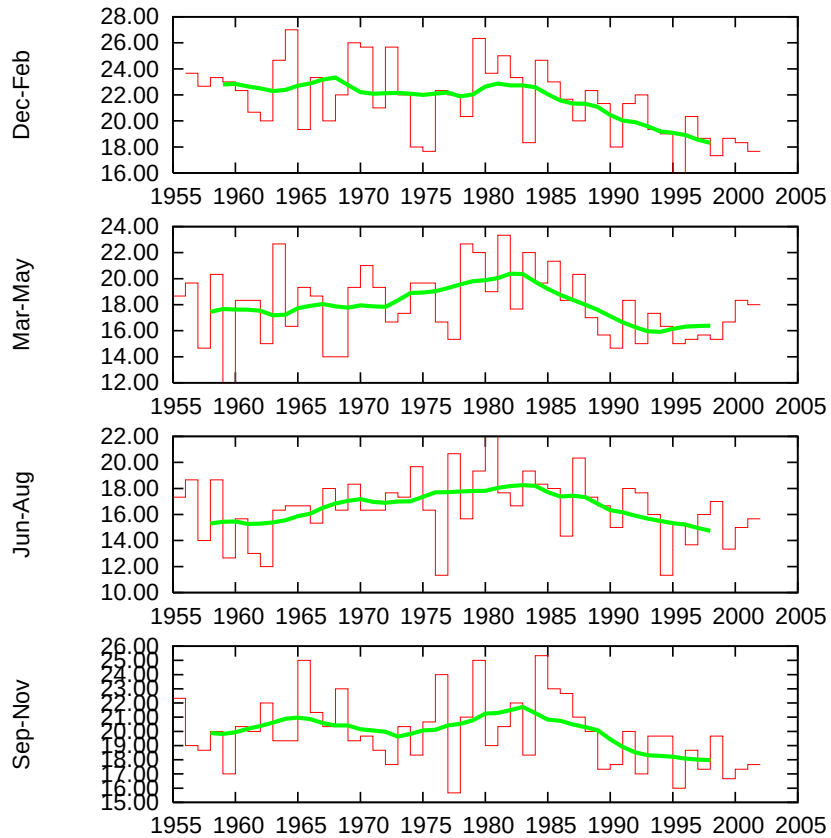


JJA

- Aerosol dip around 1985 in observations and model?
- Linear trend 1971-2007 in model and observations?

Aerosol effects





Essence sulphate column (Vertical)

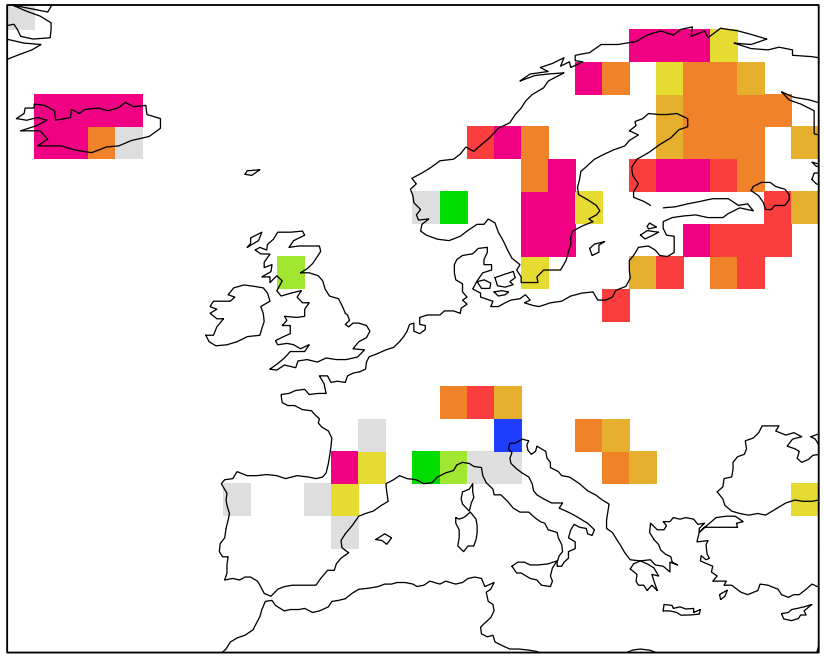
Number of days with vv<5km (Horizontal)

Aerosol effects in Holland

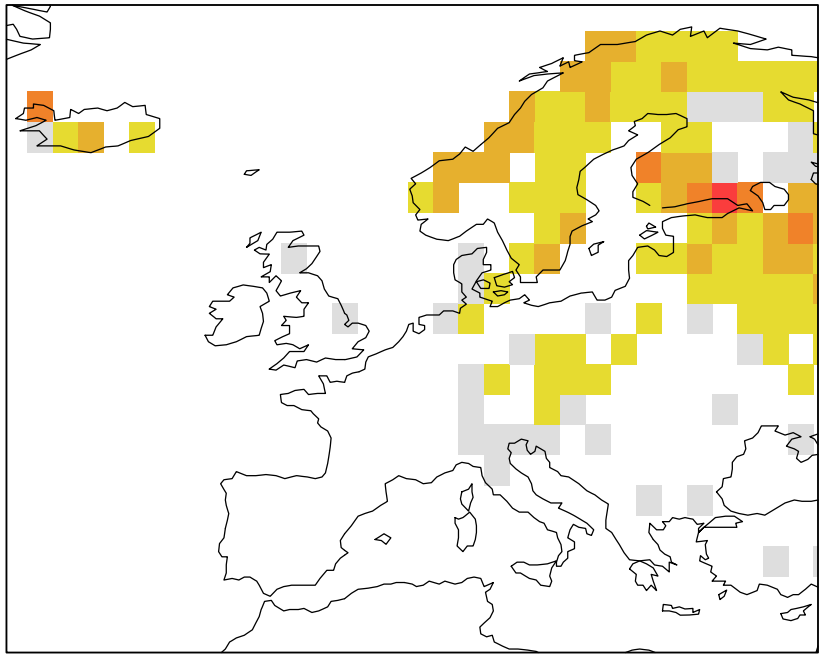
- In summer, observed aerosol concentrations in 2000 are roughly equal to those in 1950, with a peak in the early 1980s.
- In the model, summer concentrations are much lower in 2000 than in 1950.
- Hence, the model should *overestimate* the effect on the trend 1950-2000.
- However, comparing radiation measurements, it *underestimates* the effect of aerosols.

We estimate that the model underestimates the summer effects of aerosols by a factor 2 to 3. Projected on $T^{(3)}_{\text{global}}$, this translates to about 20% of the discrepancy in trends.

Snow melting around the Baltic

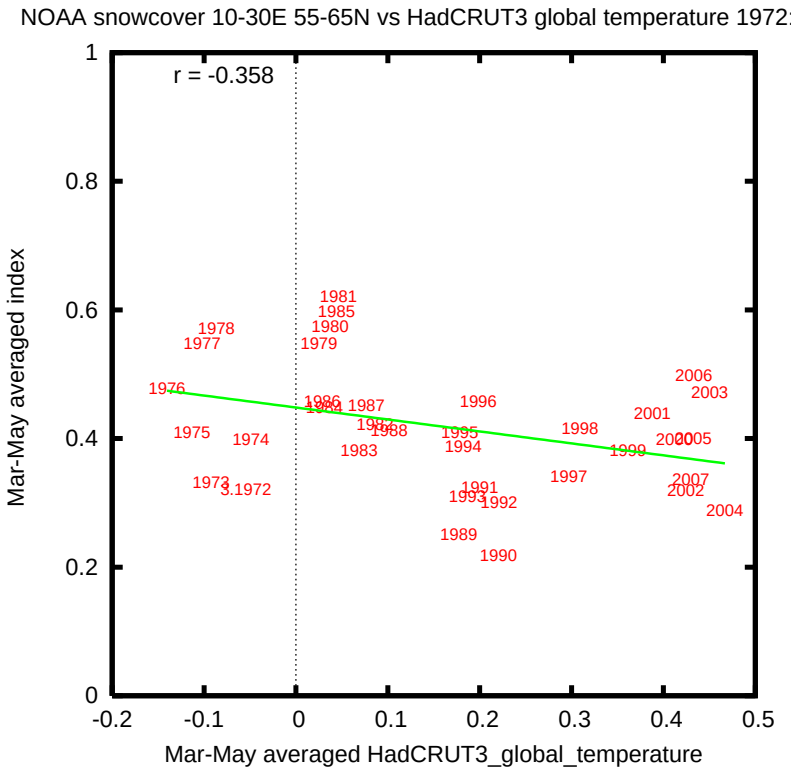


NOAA snow cover

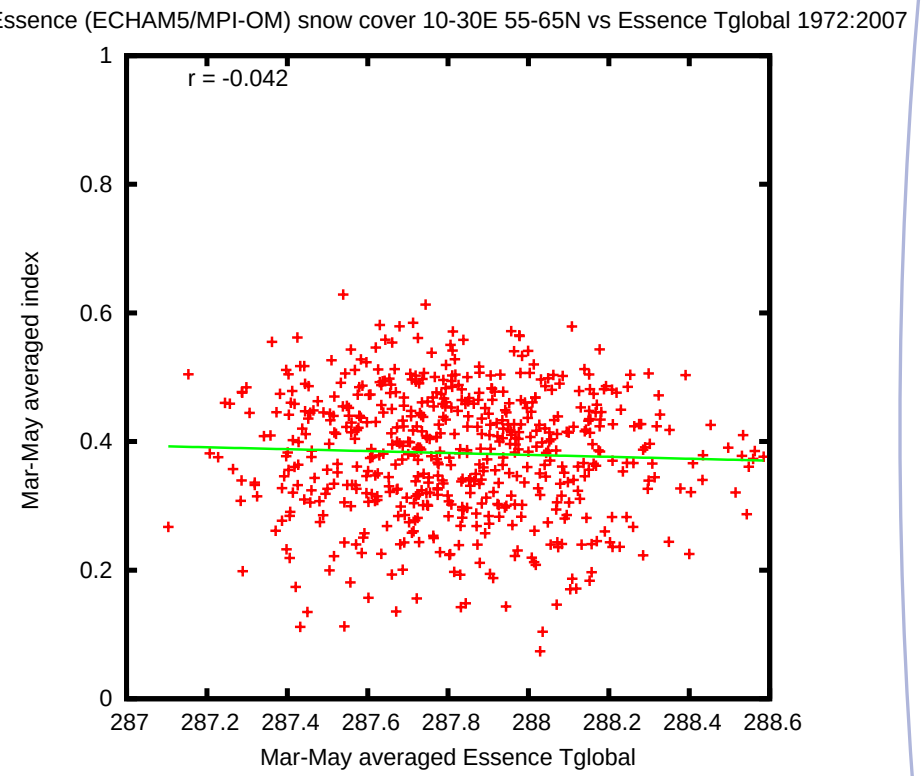


ECHAM5/MPI-OM

Snow melting around the Baltic



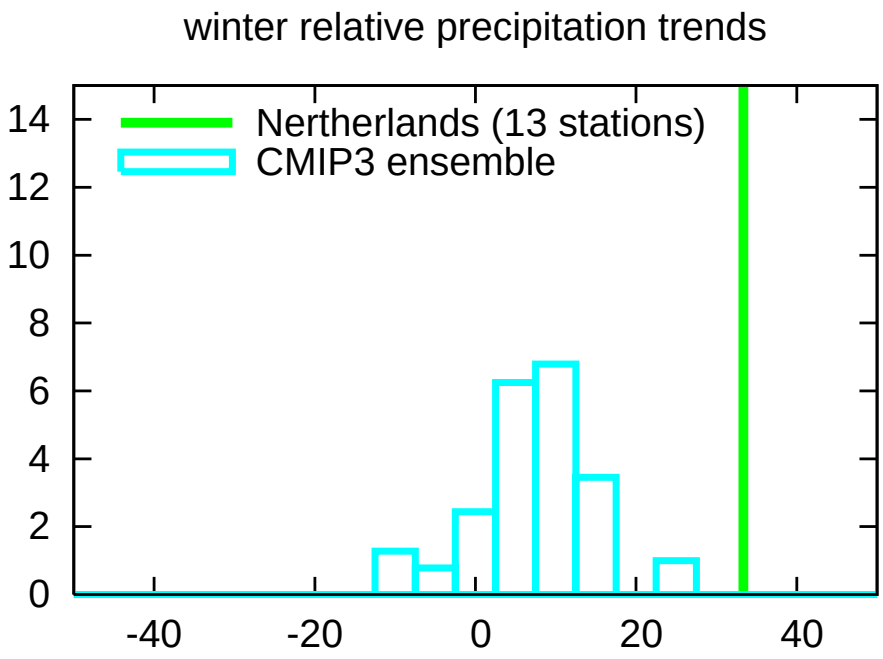
NOAA snow cover



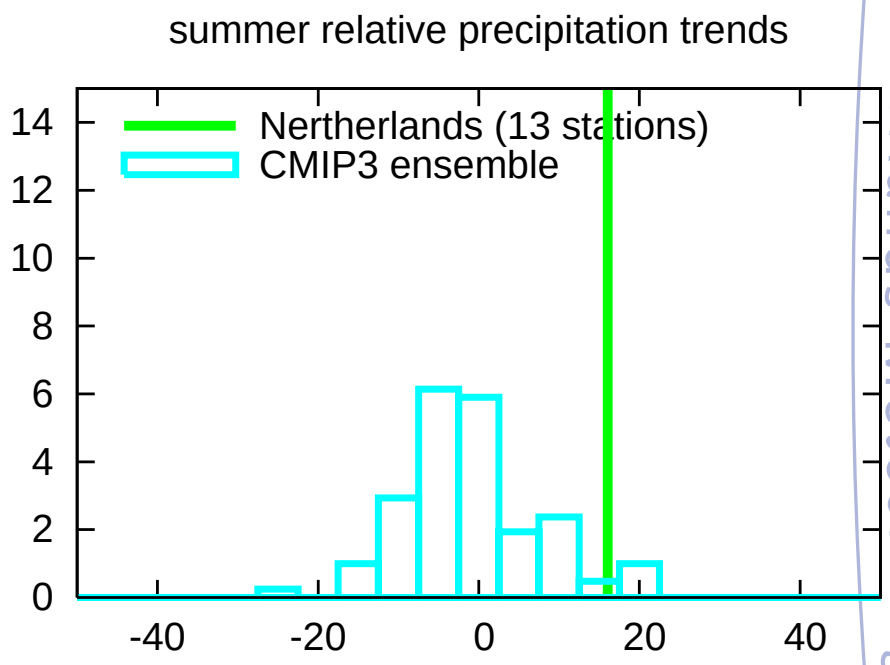
ECHAM5/MPI-OM

The difference is about 95% significant (taking a 3-yr autocorrelation into account).

Relative precipitation trends 1901-2007 in the CMIP3 ensemble



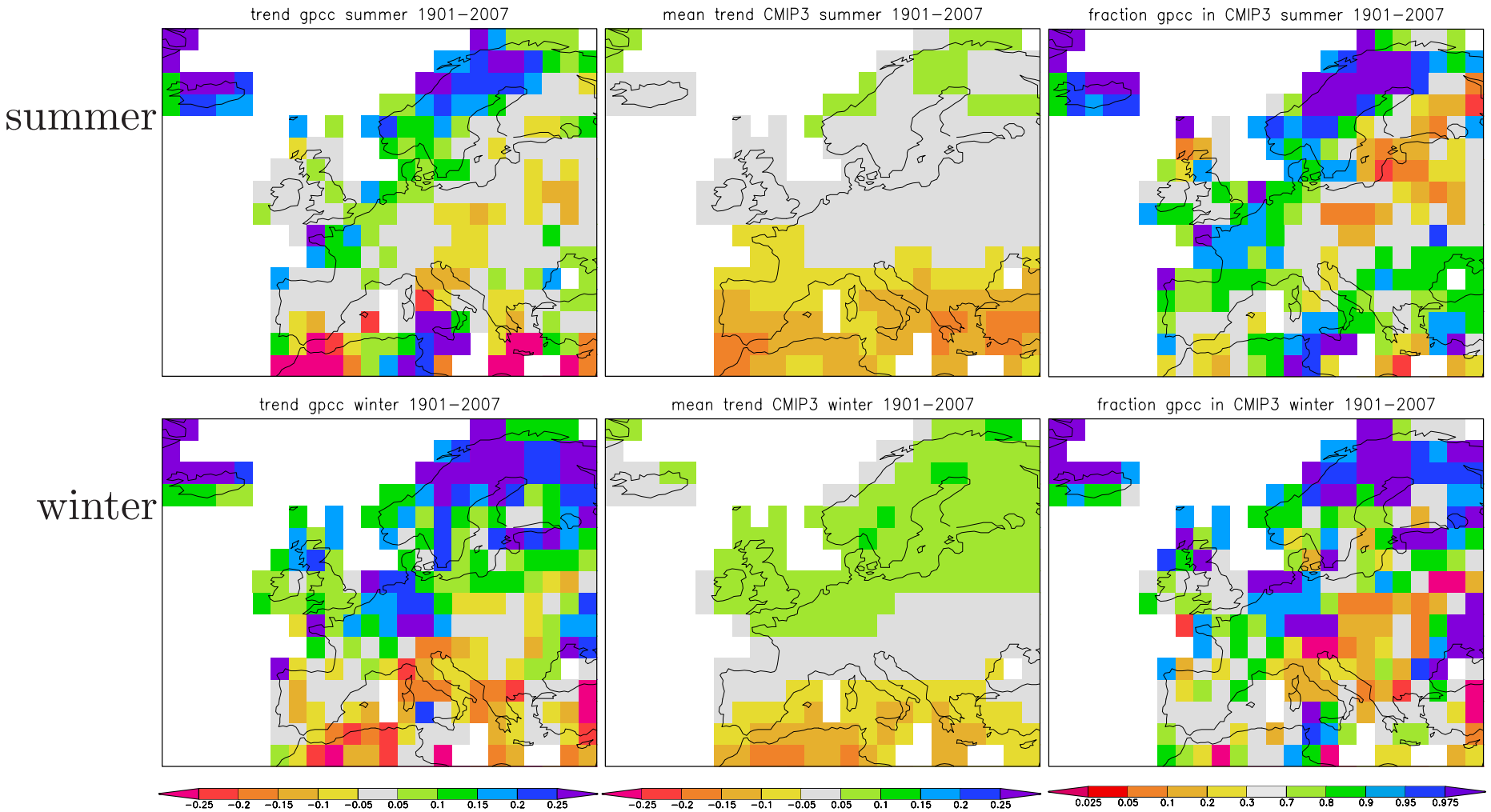
October–March [%/K]



April–September [%/K]

Multiple ensemble members of the same model have been weighed by $1/N_{\text{ens}}$ so that each model contributes equally to the histogram.

Comparison of observed trend with CMIP3



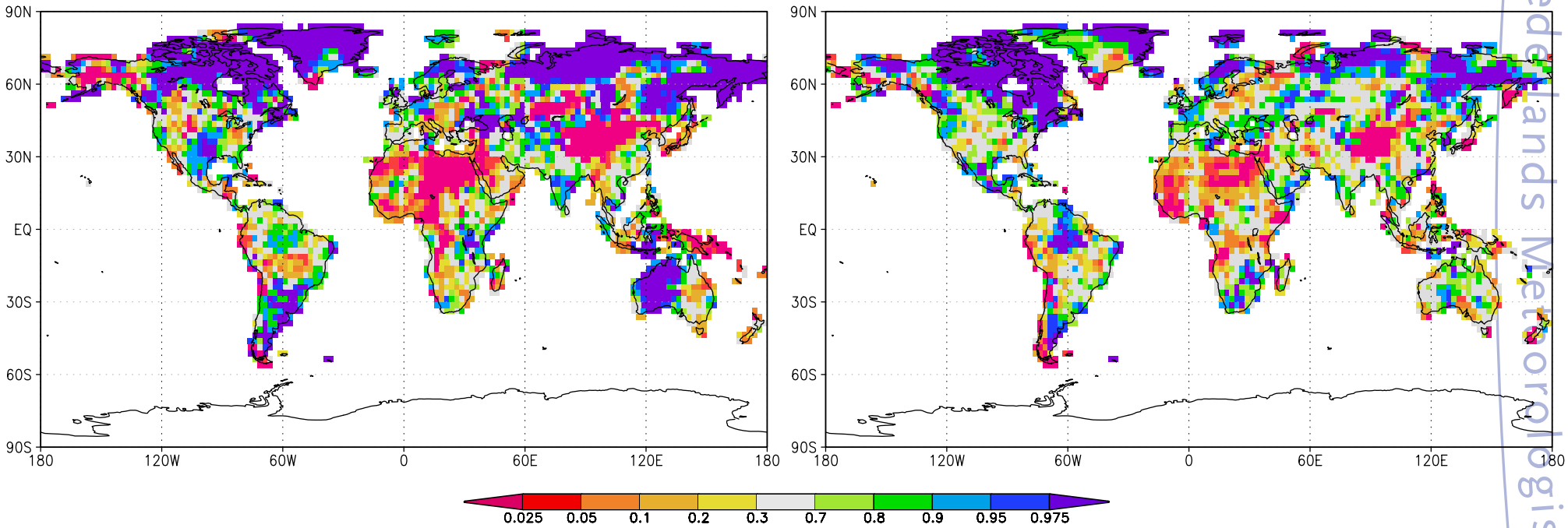
Observed rel. trend (GPCC) [1/K]

Mean modelled trend (CMIP3) [1/K]

Quantile of obs in model PDF

Also noted by Zhang et al, Nature, 2007.

Comparison of observed trend with CMIP3



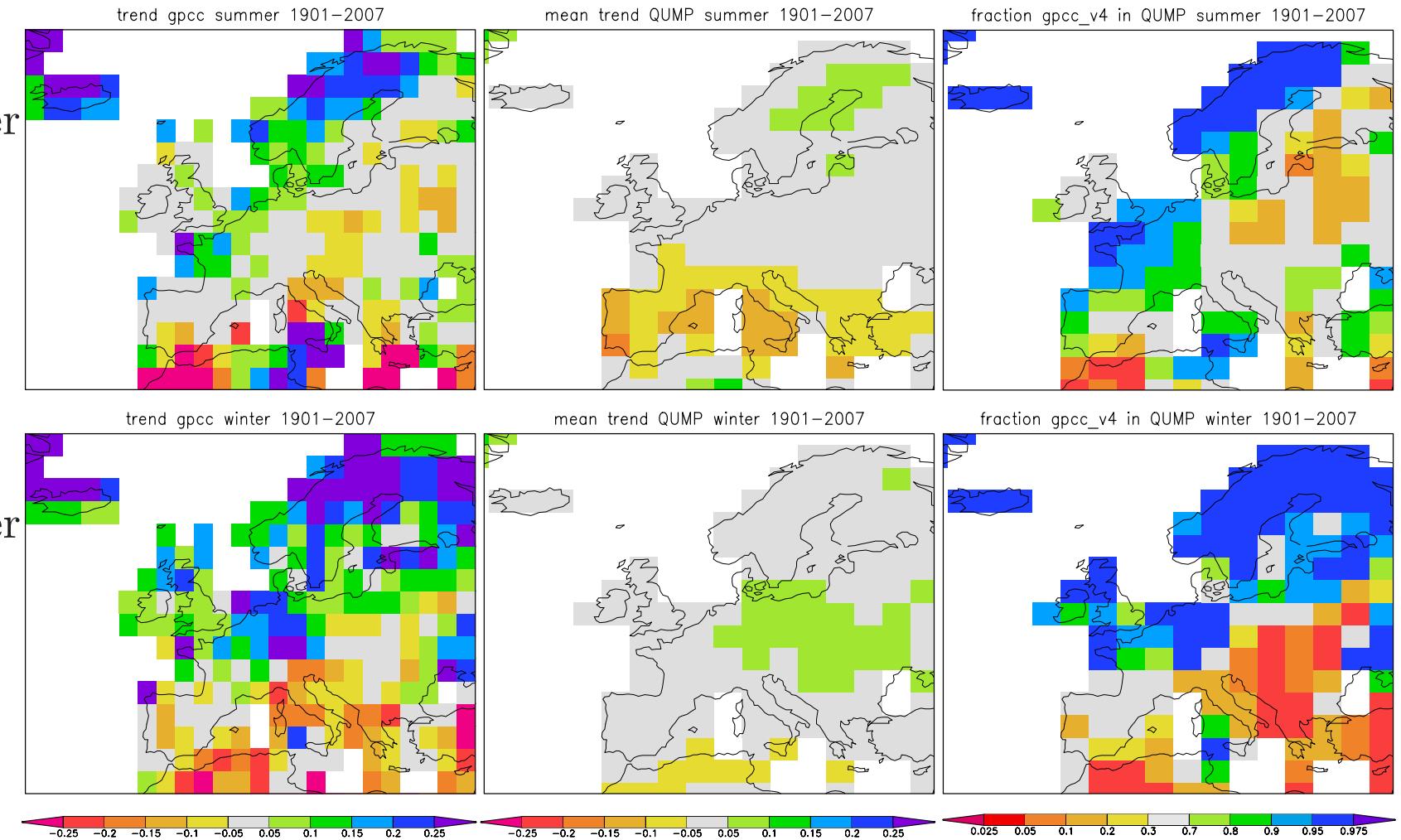
October–March

April–September

Comparison with a perturbed physics ensemble

summer

winter



Observed rel trend (GPCC) [1/K]

Mean modelled trend (QUMP) [1/K]

Quantile of obs in model PDF

Uninitialised climate simulations: conclusions 1

- Global warming is now strong enough to start verifying climate models on the temperature and precipitation trends.
- Although global and continental-scale warming are simulated well by climate models, there are large biases on regional scales.
- In coastal Europe, the warming is underestimated by a factor 1.5 to 2 so far (up to 3σ): very unlikely only due to chance weather and decadal climate fluctuations.
- In northern Europe, the trend towards more precipitation in winter is underestimated by a factor two or more.
- These biases are not caused by the coarse resolution: RCMs show the same biases.

Uninitialised climate simulations: conclusions 2

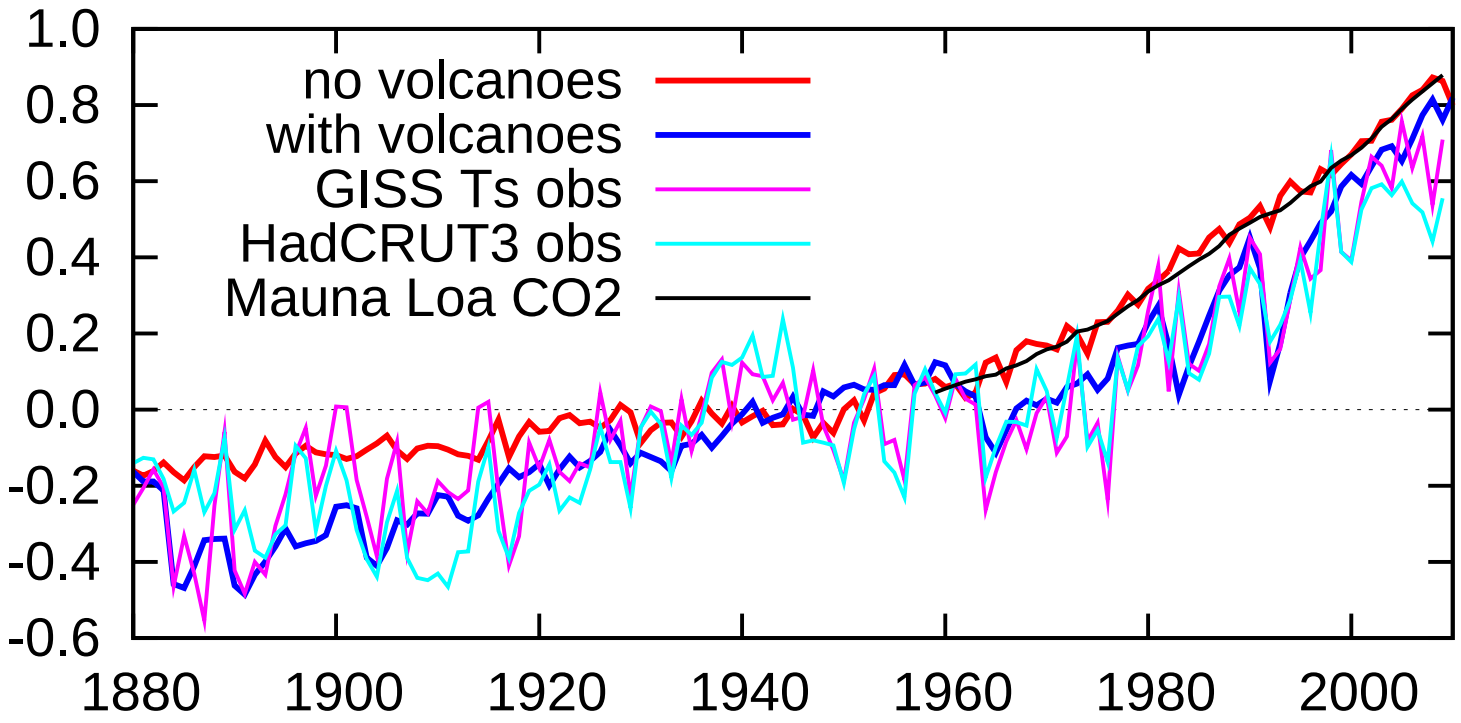
- In winter and early spring the largest changes are in the circulation, with sea-level pressure in the Mediterranean rising much faster than modelled.
- A wrong circulation in the North Atlantic Ocean in most climate models causes an underestimation of the warming trends along the Atlantic coast.
- In late spring and summer, the observed trend in short-wave radiation in the Netherlands is simulated much weaker by the ECHAM5 model. The decrease in aerosols only explain a small part of the trend 1950-2007, but seems underestimated by the models.
- In spring, snow melt trends are much smaller than observed in the Baltic

Initialised climate simulations

- Up to now we only considered the skill from boundary conditions: increasing GHG concentrations, volcanic aerosols, tropospheric aerosols.
- How do we fare when including the initial state: mainly the ocean, maybe cryosphere, land surface.

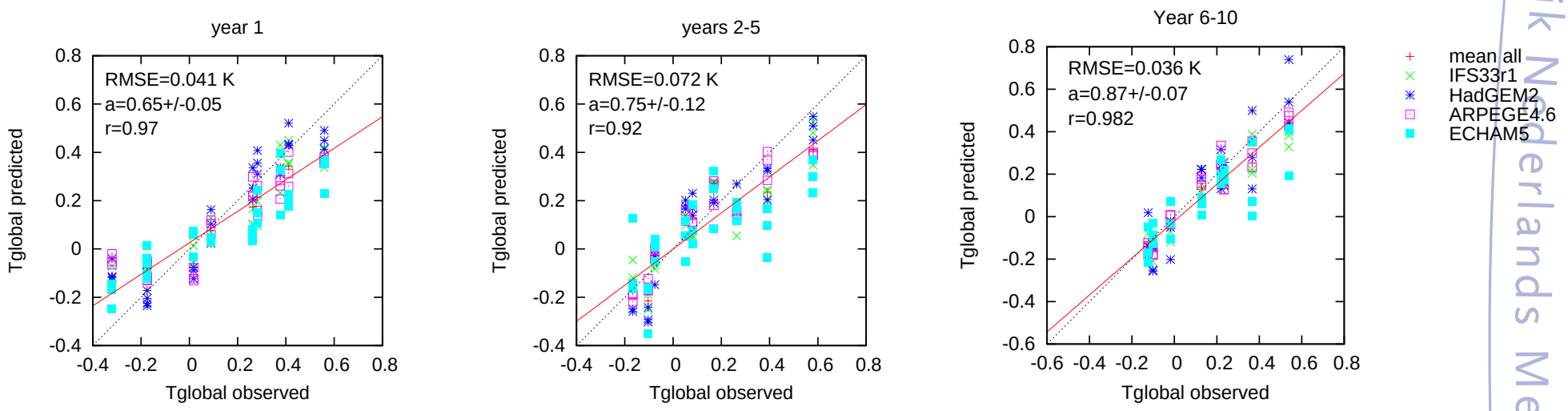
Here we look at the skill in four ENSEMBLES models (IFS33r1, HadGEM2, ARPEGE4.6, ECHAM5) that performed 10-yr hindcasts starting Nov 1 1960, 1965, 1970, . . . , 2005.

Global mean temperature



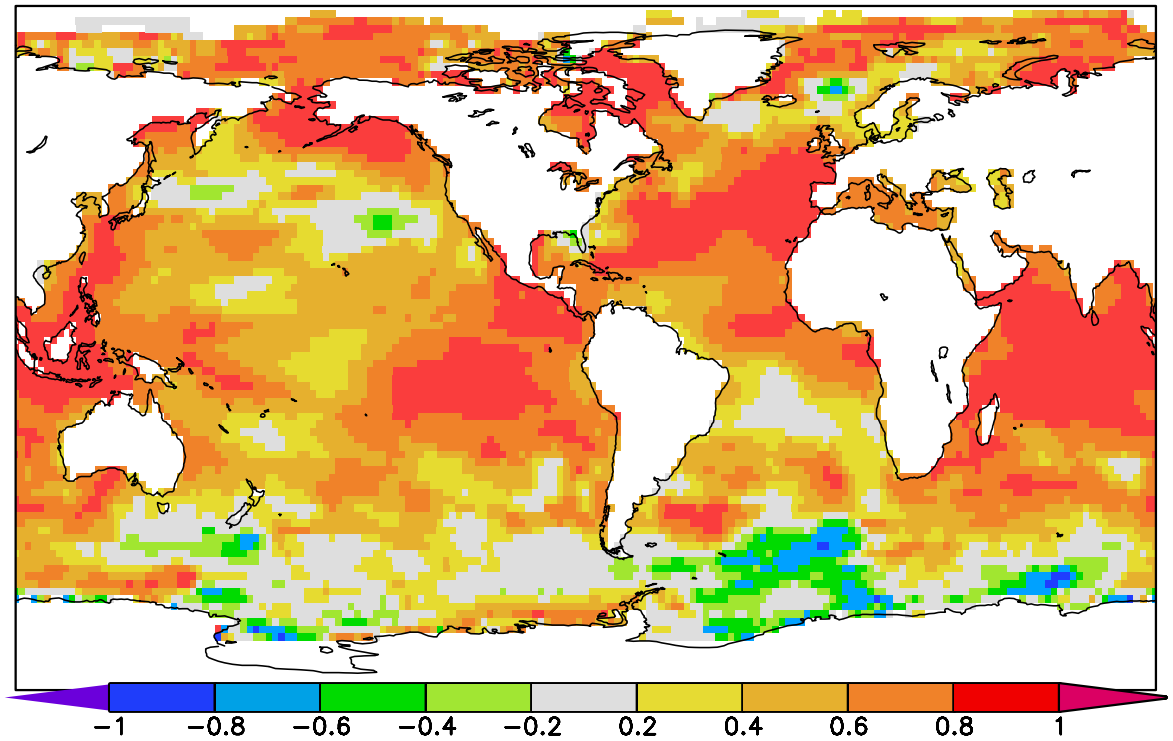
- Without volcanoes, T_{global} follows the CO_2 concentration
- The largest deviations over 1958–2008 are due to volcanoes and internal variability. Part of the latter may be predictable.

Global mean temperature forecasts



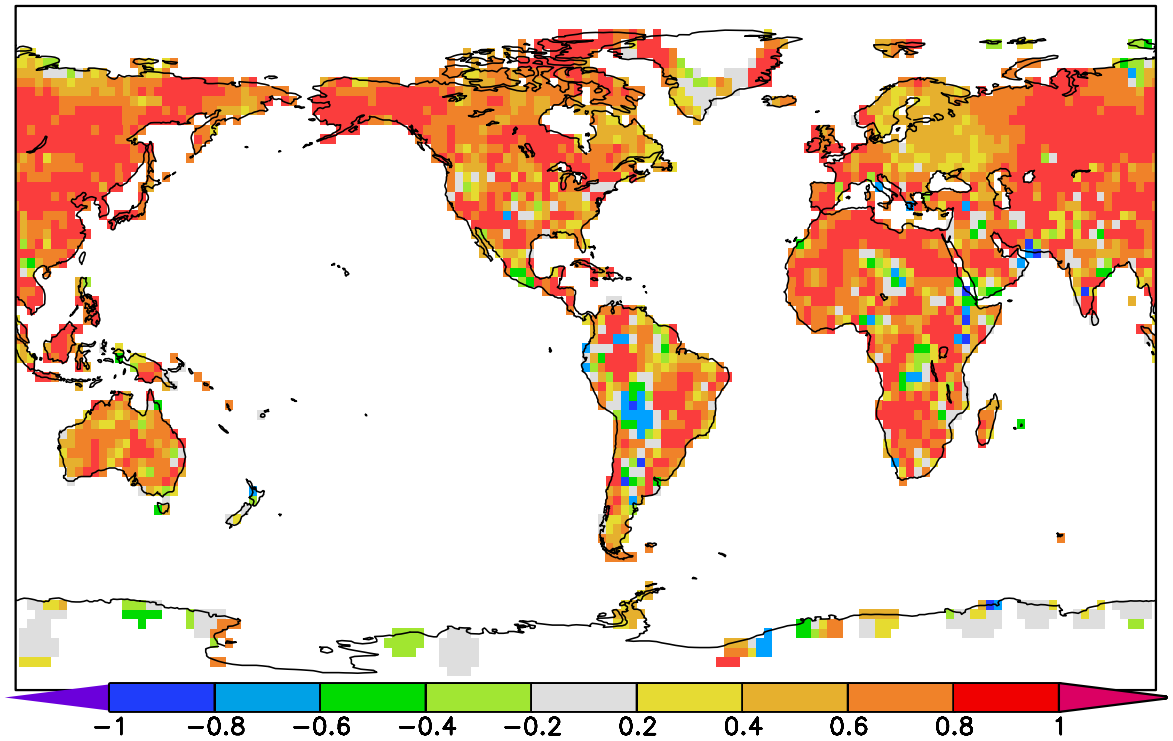
- The ENSEMBLES decadal runs have good skill in T_{global}
- Mainly due to the trend
- The CMIP3 runs without volcanoes have trend 0.80 ± 0.06 against observed T_{global} for single years, 0.94 ± 0.04 for 3-yr means, 0.99 ± 0.09 for 5-yr means. The trends in the decadal runs are 10% lower. Bias in trend / trend in bias.

Correlation skill SST globally



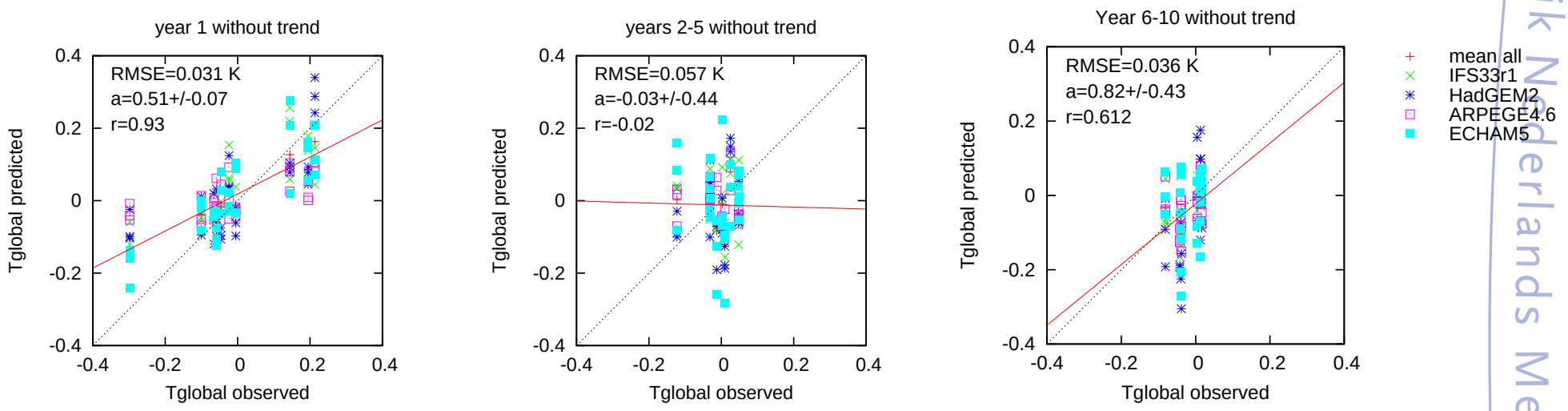
- The ENSEMBLES ensemble has good skill in SST almost everywhere (exception: areas of doubtful observations in NCDC ERSST v3b).

Land 2m temperature globally



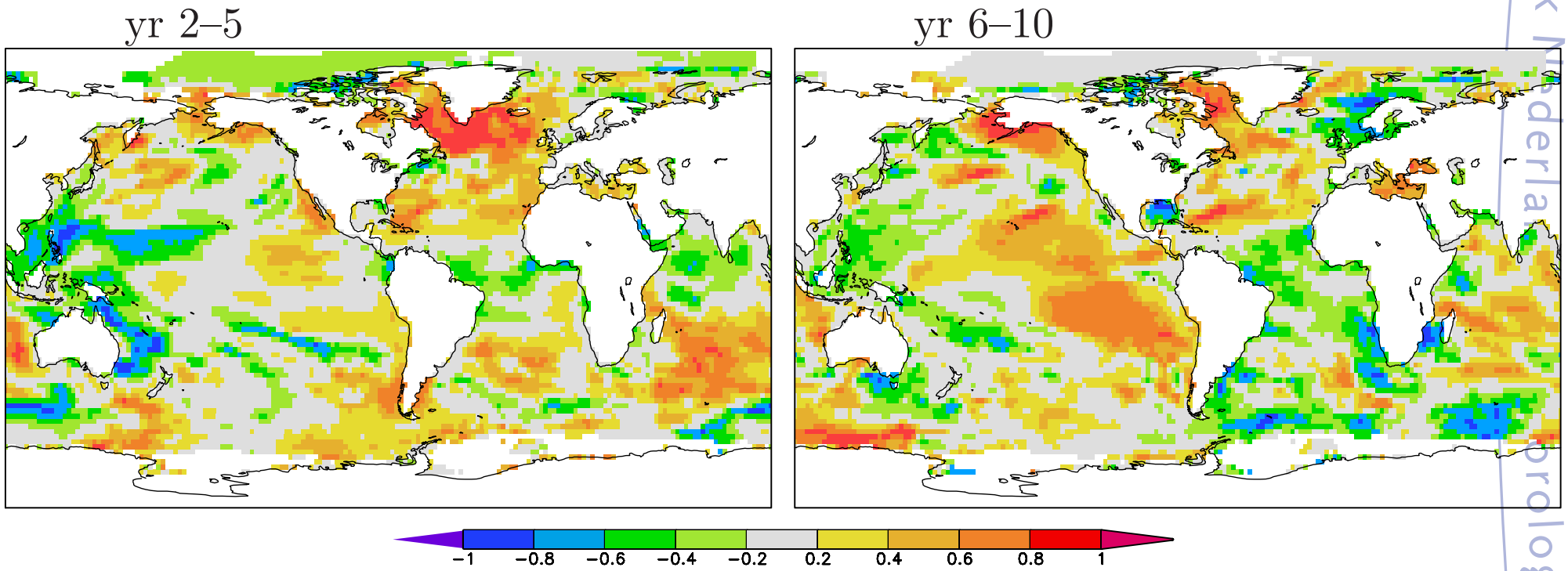
- The ENSEMBLES ensemble has good skill in T2m almost everywhere (exception: areas of doubtful observations in NCEP GHCN/CAMS).

Global mean temperature forecasts minus trend



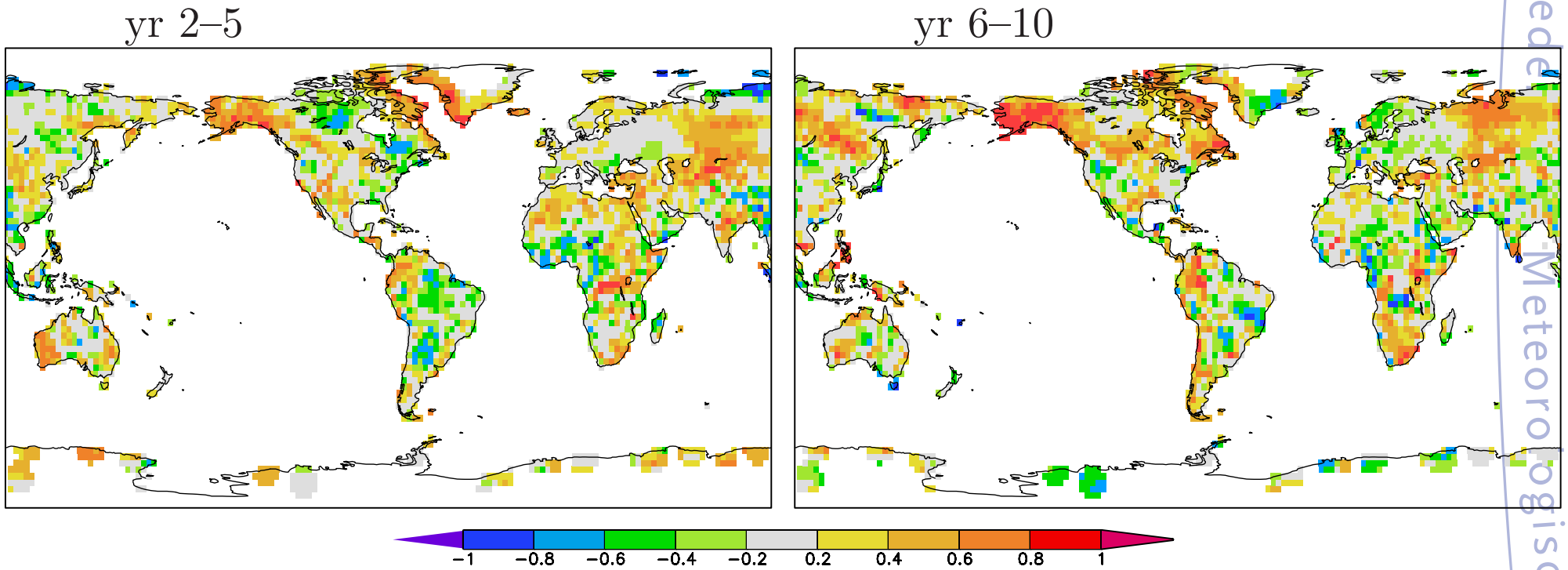
- Define the trend as projection on the CO_2 concentration
- Good skill in year 1 mainly due to persistence and ENSO (forecasts start Nov 1).
- No skill in yr 2–5 ($r = -0.01$)
- Non-significant skill in yr 6–10 ($r = 0.29, p = 0.3$)
- (Note: not yet cross-validated)

SST globally minus trend



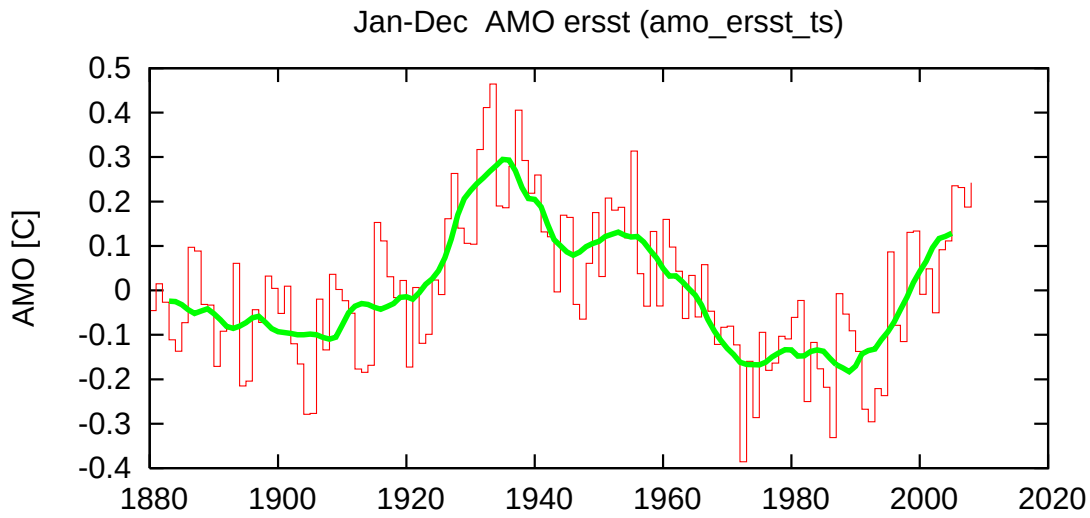
- Good skill in North Atlantic yr 2–5, some left in yr 6–10.
- Reasonable skill in decadal ENSO region in yr 6–10, maybe PDO. Why re-emergence?

Land 2m temperature minus trend



- Much less skill over land: maybe Alaska, Central Asia

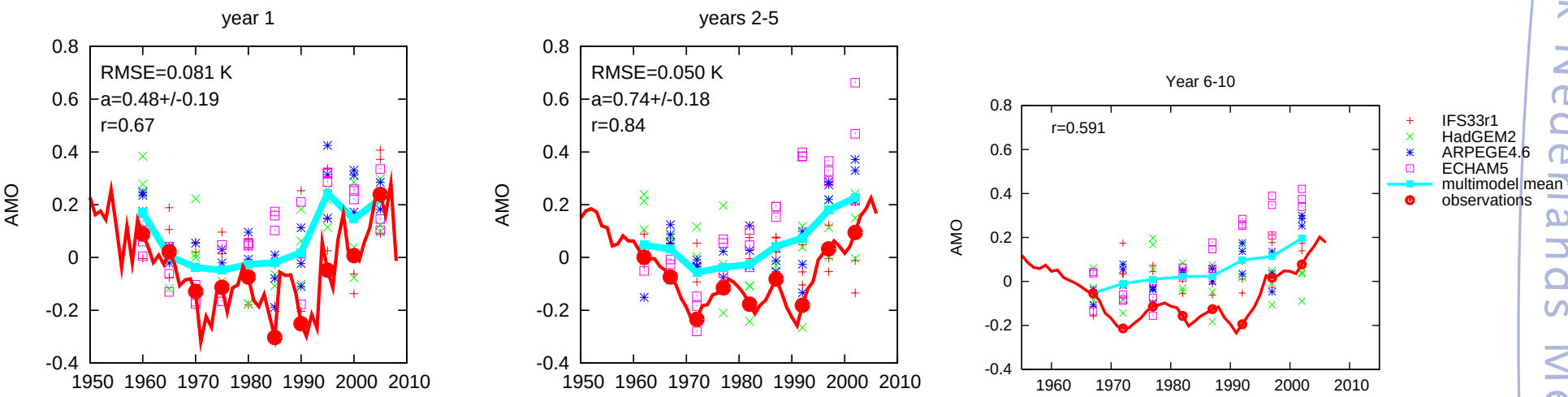
Atlantic Multidecadal Oscillation



SST EQ–60°N, 80°– 0°W
minus SST 60°S–60°N
(Trenberth and Shea 2006)

- Big shifts around 1995, 1970 and 1930, maybe 2010.
- Almost orthogonal to global warming over 1880–now and 1960–now (also in ensembles ECHAM5/MPI-OM and CCSM 3.0 relative to ensemble mean)

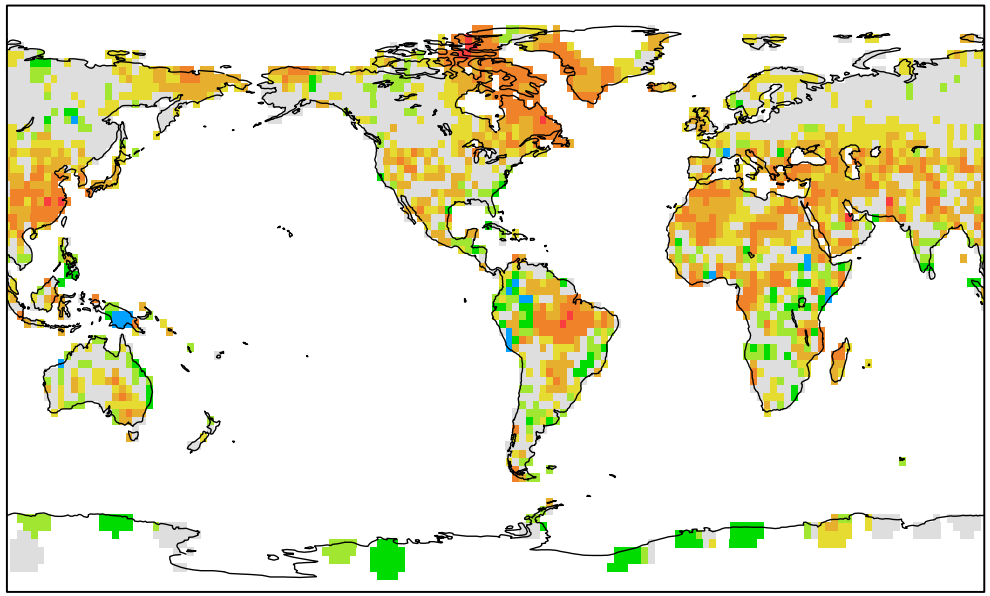
AMO forecasts



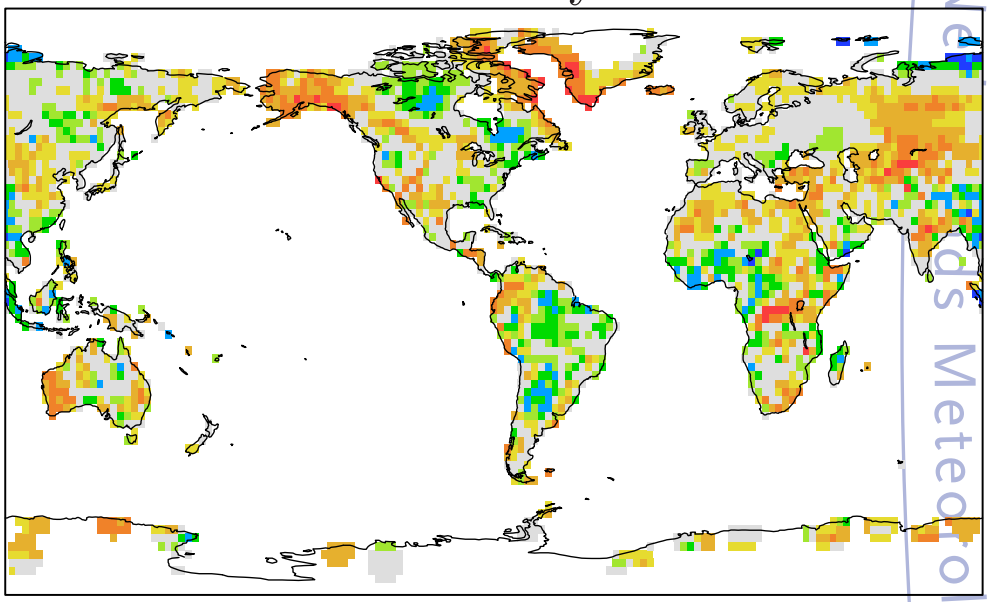
- The ENSEMBLES decadal runs show good skill in forecasting the AMO variations: drop after 1960, rise around 1995.
- In yr 2–5 $r = 0.84$, in yr 6–10 still $r = 0.59$.
- They do not reproduce the shift around 1995 except in year 1.

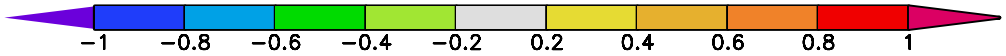
AMO teleconnections: T2m (without trend)

AMO teleconnection



Multimodel skill yr 2–5

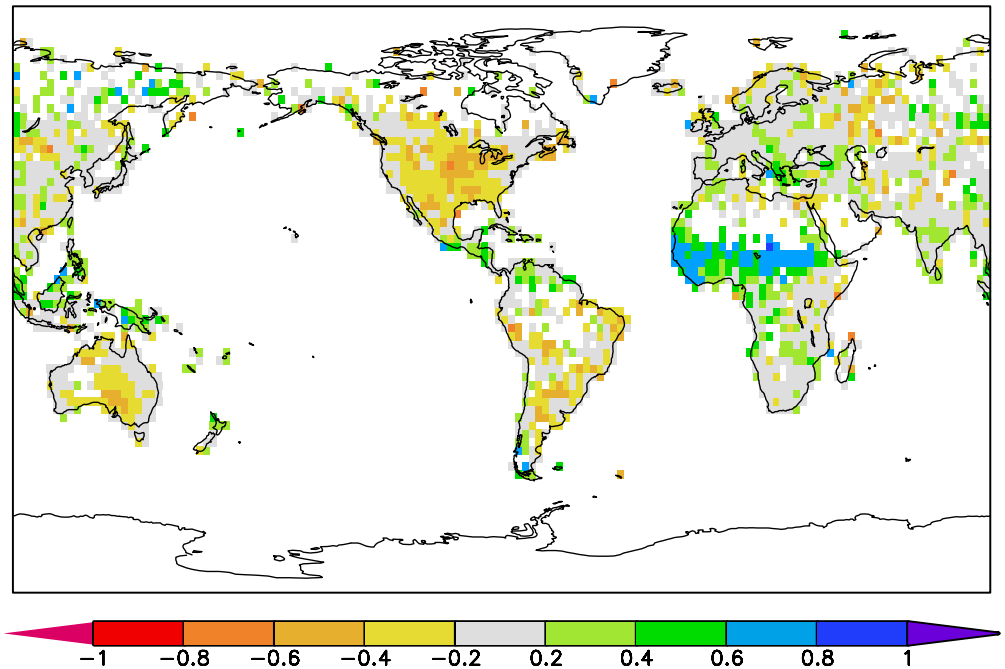




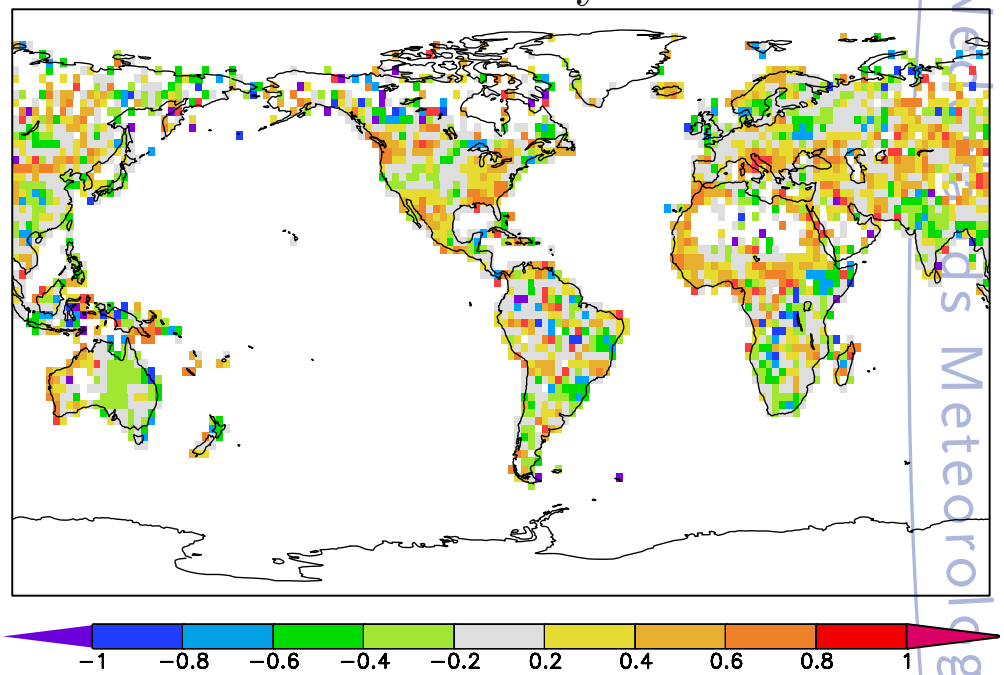
- Teleconnections are somewhat similar to skill ENSEMBLES runs: regions with a positive or negative teleconnection have overlap with the regions with positive skill

AMO teleconnections: precipitation

AMO teleconnection



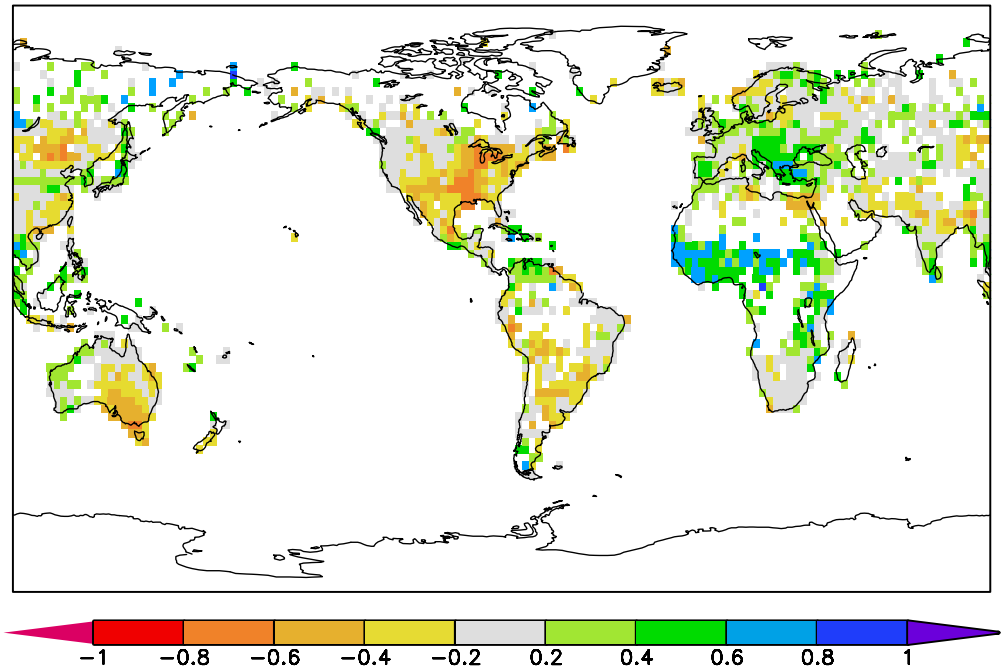
Multimodel skill yr 2–5



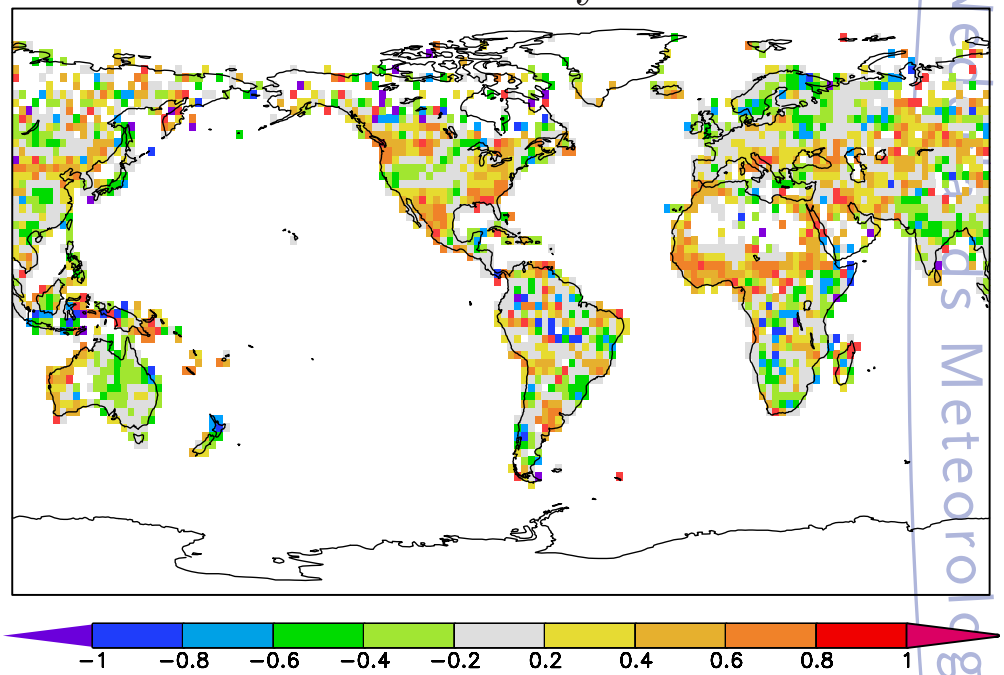
- Teleconnections are somewhat similar to skill ENSEMBLES runs: regions with a positive or negative teleconnection have overlap with the regions with positive skill

AMO teleconnections: precipitation (w/o trend)

AMO teleconnection



Multimodel skill yr 2–5



- Teleconnections are somewhat similar to skill ENSEMBLES runs: regions with a positive or negative teleconnection have overlap with the regions with positive skill

Conclusions

- In temperature most skill comes from the boundary forcing: global warming.
- The fluctuations around the trend of the global mean temperature are not predicted well by the ENSEMBLES decadal runs (IFS, ARPEGE, ECHAM5, HadGEM2).
- The models have good skill in hindcasts of the low-frequency behaviour of SST in the North Atlantic around the trend.
- The skill in temperature and precipitation after taking the trend into account overlaps with AMO teleconnections.
- Hence the ocean initial state contributes significantly to skill in these regions.