Technical Report

Department of Computer Science
and Engineering
University of Minnesota
4-192 Keller Hall
200 Union Street SE
Minneapolis, MN 55455-0159 USA

TR 18-002

A Cautionary Note on Decadal Sea Level Pressure Predictions from GCMs

Stefan Liess, Peter K. Snyder, Arjun Kumar, Vipin Kumar

February 6, 2018

A Cautionary Note on Decadal Sea Level Pressure Predictions from GCMs

Stefan Liess[1], Peter K. Snyder[2], Arjun Kumar[3], and Vipin Kumar[3]

[1] Department of Earth Sciences, University of Minnesota, Minneapolis, MN

[2] Department of Soil, Water, and Climate, University of Minnesota, St. Paul, MN

[3] Computer Science & Engineering, University of Minnesota, Minneapolis, MN

Submitted to
Advances in Climate Change Research

Corresponding author: Stefan Liess, Department of Earth Sciences, University of Minnesota, 116 Church Street SE, Minneapolis, MN 55455
Phone: (612) 624-1620, Fax: (612) 625-3819, Email: liess@umn.edu

Abstract

Decadal prediction of sea level pressure (SLP) plays an important role in regional climate prediction, because it shows changes in atmospheric behavior on time scales that are relevant for policy makers. These changes consist of a combination of externally forced and internally driven climate system characteristics.

A comparison of SLP trends in a subset of seven Coupled Model Intercomparison Project (CMIP) phase 5 general circulation models (GCM) during the satellite-era to their CMIP3 counterparts reveals an unrealistically strong forecast skill in CMIP3 models for trend predictions for 2001-2011 when using the 1979-2000 period to train the forecast.

Boreal-winter SLP trends over five high-, mid-, and low-latitude zones were calculated over a two-decade initialization period for each ensemble member and then ranked based on their performance relative to observations in all five zones over the same time period. The same method is used to rank the ensemble members during the following decade. In CMIP3, 17 out of 38 ensemble members retain their rank in the 2001-2011 hindcast period and 3 retain the neighboring rank. However, these numbers are much lower in more recent CMIP5 decadal predictions over a similar period with the same number of ensembles. The conclusion to consider the forecast skill in CMIP3 predictions during the 2001-2011 as unrealistic is corroborated by comparisons to earlier periods from the 1960s to the 1980s in both CMIP3 and CMIP5 simulations. Thus, although the 2001-2011 CMIP3 predictions show statistically significant forecast skill, this skill should be treated as a spurious result that is unlikely to be reproduced by newer more accurate GCMs.

1 Introduction

Observational analyses have documented distinct patterns of Atlantic ocean variability with decadal (8–12 years) and multidecadal (30–80 years) time scales (Delworth et al. 2007). General circulation models (GCMs) from the third and fifth Coupled Model Intercomparison Project (CMIP) have succeeded in capturing some aspects of this observed variability (Medhaug and Furevik 2011; Zhang and Wang 2013). The resulting decadal predictability remains influenced by the initial conditions from slow components of the climate system (Collins 2007; Pohlmann et al. 2009). Some techniques to improve decadal predictions include bias-correction of initial conditions (Laepple et al. 2008). GCM simulations show that changes in sea surface temperature (SST) and sea ice variability can be related to the Atlantic Meridional Overturning Circulation (AMOC) (Mahajan et al. 2011a; Zhang and Wang 2013), and the strength is potentially predictable for up to a decade, with skill varying among models (Pohlmann et al. 2009; Matei et al. 2012).

The AMOC contributes a substantial fraction of the low-frequency variability of the basin-averaged North Atlantic SSTs, i.e. the Atlantic Multidecadal Oscillation (AMO) (e.g., Folland et al. 1986; Delworth and Mann 2000; Dima and Lohmann 2007; Zhang and Wang 2013), and thus offers some predictability of the global and Northern Hemisphere mean temperatures on the decadal scale (Knight et al. 2005). Keenlyside et al. (2008) found increased prediction skill after initializing ECHAM5/MPI-OM with SSTs restored from previous simulations that were nudged to observed SSTs. On decadal time scales, GCM predictions of AMOC variations depend critically on the initial state of the AMOC in the modeled climate (Keenlyside et al. 2008; Pohlmann et al. 2009; Mahajan et al. 2011b; Yeager et al. 2012). An alternative hypothesis has recently linked the AMO to surface winds and related sea level pressure (SLP) variability (Clement et al. 2015).

Decadal trends in Arctic SLP have previously been noted by Walsh et al. (1996), and Latif et al. (2000) have investigated the predictability of decadal winter-time North Atlantic SLP variations. Walsh et al. (1996) found considerable evidence that these SLP variations on timescales of several years to decades may be predictable in their analysis of the ECHAM4 model, which is a predecessor to the atmospheric version of the CMIP3 model ECHAM5/MPI-OM. Latif et al. (2000) found that the simulated North Atlantic Oscillation (NAO) is significantly correlated at 0.32 with observations, and smoothing both time series with canonical correlation analysis reveals a correlation of 0.65. In general, GCMs are known to exhibit more zonally symmetric annular modes than observations (Gerber et al. 2008; Xin et al. 2008; Gerber et al. 2010). Thus, zonally averaged trends are considered appropriate for capturing the variations in simulated annular mode activity.

Although the Pacific response to AMOC variations associated with the AMO is very small (van Oldenborgh et al. 2009), it has been suggested that the AMO and the 60-year component of the Pacific Decadal Oscillation (PDO) are signatures of the same oscillation cycle (d'Orgeville and Peltier 2007), and the AMO can be related to global SST patterns including the Southern Hemisphere (Zanchettin et al. 2014). Low-frequency ocean variability can already be successfully predicted, especially over the North Atlantic and the western Pacific Oceans (Pohlmann et al. 2009; Kim et al. 2012). However, in

contrast to the large-scale findings by Latif et al. (2000) described above, DelSole et al. (2011) found that these internal multidecadal SST patterns have no obvious statistically significant correlation with higher-frequency concurrent atmospheric surface winds, SLP, or precipitation at individual grid cells, which motivates an area-averaged SLP analysis over larger regions.

Because multidecadal oscillations of geopotential height and temperature can stretch over multiple ocean basins (Lee and Hsu 2013), it is tempting to analyze if large-scale SLP changes over these larger zones can be predicted on the decadal scale, similar to the previously discussed results by Latif et al. (2000). The dependence of annual mean SLP on latitudinal zones and seasonal variation has been previously analyzed by Trenberth (1981, his table 8). Correlations between northern and southern annular modes are linked to Madden-Julian oscillation (MJO) activity (Flatau and Kim 2013), with lags in daily data being smoothed in the monthly averages used in this study.

Gillett et al. (2003; 2005) and Miller et al. (2006) have associated a decrease in high-latitude SLP with anthropogenically influenced increase in greenhouse gases (GHG) and ozone depleting substances with the latter being primarily important in the Southern Hemisphere forcing (Gillett and Thompson 2003; Arblaster and Meehl 2006; Wilmes et al. 2012; Gillett and Fyfe 2013; Simmonds 2015). This negative SLP trend over high latitudes is accompanied by an SLP increase over low and mid-latitude regions, resembling the annular modes. Superimposed on these long-term externally forced trends are low frequencies in the annular modes that can provide varying SLP trends on the decadal time scale. The combination of externally forced variability and the internal low-frequency variability that comprises the decadal climate time scale is the focus of this study.

The Northern Annular Mode (NAM) exhibits the strongest variability during the December-January-February (DJF) season (Hurrell 1995, 1996; Gillett and Fyfe 2013), whereas the Southern Annular Mode (SAM) variability remains similar for all seasons (Trenberth 1991) but also shows a strong axisymmetric pattern around the South Pole in DJF (Simmonds 2015). It is suggested that the NAM is not only closely related to the NAO (Wallace 2000; Feldstein and Franzke 2006), but is also related to the Pacific North America pattern (Thompson and Wallace 2001) and other SLP patterns over Eurasia (Philipp et al. 2007). These large-scale relations can be well represented in GCMs (Stoner et al. 2009; Casado and Pastor 2012; Handorf and Dethloff 2012; Pastor and Casado 2012), thus justifying the circumglobal approach in this study. Circumglobal Rossby wave trains in the Northern Hemisphere have been identified for boreal summer (Ding and Wang 2005; Schubert et al. 2011; Teng et al. 2013) and winter (Branstator 2002; Liess et al. 2017). Here we analyze the global SLP field because it is closely linked to near-surface winds and other climate processes that are relevant for policy makers, although it is noisier than mid- and upper- tropospheric geopotential height fields. SLP trends are important indicators for changing weather and climate patterns (Gillett et al. 2013). The role of stratospheric and other non-oceanic drivers in decadal climate predictability has been discussed in detail by Bellucci et al. (2015).

The time frame since 1979, when satellite data became widely available, or subsets of time frames within this recent period have been widely used for observational studies (Trenberth et al. 2005) and CMIP3 model intercomparison studies (e.g.; Pincus et al. 2008; Karpechko et al. 2009). Although external forcing is similar for all 20th century

CMIP3 simulations, internal AMO and PDO phases have been found to be quite different (Meehl et al. 2009). Even the more recent decadal CMIP5 simulations with observed initial conditions can produce a variety of AMO and PDO phases, and spatial AMO representation shows both improvements and deterioration from CMIP3 to CMIP5 depending on the model (Ruiz-Barradas et al. 2013). However, in general, AMOC and AMO simulations have been improved in CMIP5 compared to CMIP3 (Zhang and Wang 2013).

Section 2 of this paper introduces a simple method that (a) ranks CMIP3 and CMIP5 ensemble members by their representation of different phases within the multidecadal SLP variability and (b) attempts to improve predictions of large-scale SLP trends. Results of the analysis are described in section 3. Section 4 provides a summary and a conclusion.

2 Methods

We calculate linear trends of SLP for a two-decade spinup period (from now on referred to as initialization period) separately for high-, mid-, and low-latitude zones, and rank model simulations by their ability to match observed trends. For a limited subset of CMIP3 GCMs, we also show predictions that seem to be improved during the 2001-2011 decade (from now on referred to as hindcast period) when excluding members from the multimodel ensemble (MME) that would produce initial conditions from opposing phases during the prior spin-up period.

This study uses seasonal mean HadSLP2 observations (Allan and Ansell 2006) for the DJF season to determine relationships between decadal trends for the five high (80°S-50°S and 50°N-80°N), mid (50°S-20°S and 20°N-50°N), and low (20°S-20°N) latitude zones as identified by their deviation from the average SLP. Fig. 1 shows the average SLP during DJF for 1979-2000. Also shown on the right is the zonal mean value in comparison to the 1013.25 hPa global mean value based on the U.S. standard atmosphere (NASA 1976), which shows that the five regions mentioned above can be distinguished by their deviation from the global mean. The southernmost region shows a strong negative deviation, the two midlatitudinal zones show moderate positive deviations, and the northernmost and tropical zones show virtually no deviations. The long-term trend pattern shown by Gillett and Stott (2009, their Fig. 1a) corroborates these distinctions and shows that long-term changes enhance most of the deviations from the global mean. However, previous research by Walsh et al. (1996) and Latif et al. (2000) shows that there is decadal variability within this long-term trend.

Decadal climate variability in mid- and upper latitudes is reflected in modifications of the southern and northern annular modes. Thus, the southernmost two zones in our study correspond roughly to the two ends of the SAM, and we select the northernmost two zones to capture trends in the NAM.

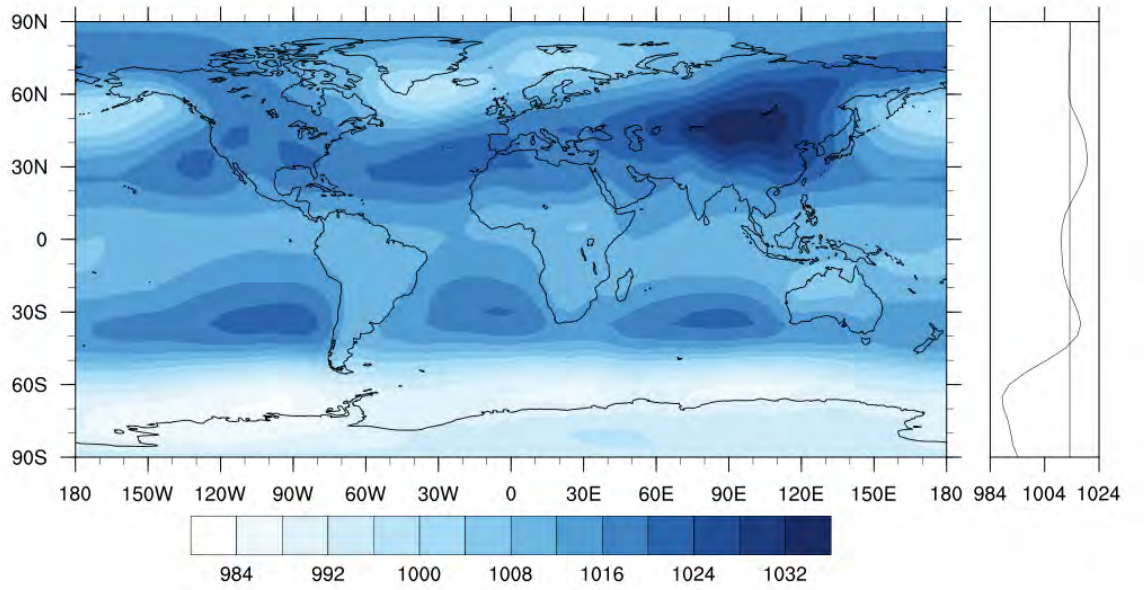

Fig. 1: Global SLP field during DJF for 1979-2000 from HadSLP2 observations. The right panel shows zonal mean values in comparison to the global mean value of 1013.25 hPa (vertical line).

These northern high and mid latitude zones also form the border between the subpolar and subtropical gyre in the North Atlantic, which are connected by a heat content dipole (Yeager et al. 2012). Poleward propagating Rossby waves can provide symmetric patterns around the equator (L'Heureux and Thompson 2006; Flatau and Kim 2013; Liess et al. 2014), prompting a combined analysis of SLP trends over both hemispheres. SLP trends are statistically significant primarily over low latitudes where monthly variability is relatively low (Gillett and Stott 2009). However, Gillett et al. (2003; 2005) also state that while reanalysis data and observations reflect this trend in the SLP gradient since the 1950's, both positive and negative trends are up to one order of magnitude lower in GCM ensemble simulations, mainly because simulations with opposing phases of large-scale oscillations are averaged to form the ensemble mean. Thus, although the observed trend is a combination of trends from internal and external forcings, ensemble simulations that smooth out the internal variability indicate the portion of external variability. Furthermore, it is assumed here that individual GCM simulations with trends similar to observations also exhibit a similar combination of trends from internal and external forcings. Systematic model errors are thus neglected in this study.

Here, we analyze 38 individual ensemble members from nine CMIP3 coupled atmosphere-ocean GCMs. Each GCM provides at least three ensemble members with continuous simulations for the 20th century as well as the A1B (Nakicenovic et al. 2000) scenario after 2000. Ensembles for each GCM use slightly different initial conditions and are thus different realizations of simulated climate characteristics. We compare these 38 CMIP3 ensemble members to 38 ensemble members from CMIP5 decadal simulations which are successors of the CMIP3 models, where available. The CMIP5 models use observed external forcing based on the RCP4.5 scenario (van Vuuren et al. 2011) after 2005.

The CMIP3 and CMIP5 GCMs as well as the number of ensembles for each GCM are listed in Table 1. However, it should be noted that although many studies compared CMIP3 results to observations starting with the satellite era in 1979, decadal CMIP5 simulations started at the first year of every decade, in this case in 1981. In order to compensate for this discrepancy, our study utilizes different time frames for CMIP3 and CMIP5 simulations that remain within 10% of each other's time frames and overlap over 90% of the time. The CMIP3 analysis uses the 1979-2000 initialization and 2001-2011 hindcast periods, whereas the CMIP5 analysis comprises of the 1981-2000 initialization and 2001-2010 hindcast periods. Thus, the length of the hindcast period is exactly 50% of the initialization period in both analyses. The time period is cut in half from the initialization period to the hindcast experiments, because forecasting skills of atmospheric oscillations are often less than one cycle (i.e. information about a 22-year period might be required for an 11-year prediction), and predictability is suggested to decrease strongly after one decade (Pohlmann et al. 2009; Kim et al. 2012).

Table 1: List of GCMs used in this study with number of ensemble members in parentheses.

Center	CMIP3 GCM	Reference	CMIP5 GCM	Reference
NCAR (USA)	CCSM3 (7)	(Collins et al. 2006)	CCSM4 (10)	(Gent et al. 2011)
CCCMA (Canada)	CGCM3.1(T47) (5)	(McFarlane et al. 2005)	CanCM4 (10)	(Chylek et al. 2011)
MPI (Germany)	ECHAM5/MPI-OM (4)	(Roeckner et al. 2003)	MPI-ESM-LR (3)	(Raddatz et al. 2007)
LASG (China)	FGOALS-g1.0 (3)	(Yu et al. 2011)	FGOALS-g2 (3)	(Li et al. 2013)
NOAA-GFDL (USA)	GFDL-CM2.1 (3)	(Delworth et al. 2006)		
NASA-GISS (USA)	GISS-EH (3)	(Schmidt et al. 2006)		
NASA-GISS (USA)	GISS-ER (5)	(Schmidt et al. 2006)		
MIROC (Japan)	MIROC3.2(medres) (3)	(Nozawa et al. 2007)	MIROC4h (3)	(Sakamoto et al. 2012)
MIROC (Japan)			MIROC5 (6)	(Watanabe et al. 2010)
MRI (Japan)	MRI-CGCM2.3.2 (5)	(Yukimoto et al. 2006)	MRI-CGCM3 (3)	(Yukimoto et al. 2012)

For an estimate of model skill, we analyze individual ensemble members and compare the results to observed SLP trends over the zones defined above. The trend analysis is performed with linear least squares regression, which identifies the slope or trend *T(x)* at grid point *x* that minimizes the deviations of the sums of squares from the input values:

$$T(x) = \frac{\sum_{t \in S(x)} \left(i(t,\ x) - \frac{1}{\#S(x)} \sum_{t \in S(x)} i(t',\ x) \right) \left(t - \frac{1}{\#S(x)} \sum_{t \in S(x)} t' \right)}{\sum_{t \in S(x)} \left(t - \frac{1}{\#S(x)} \sum_{t \in S(x)} t' \right)^2} \quad (1)$$

with *i* being the input value at location *x* and time step *t*, and *#S(x)* being the number of time steps in time series *S(x)*. The statistical significance of the trend is calculated from the ratio between the estimated trend and its standard error (Santer et al. 2000).

In addition to testing the significance of individual trends, we test if individual ensemble members of a given GCM retain large-scale information from the initialization

period for trends during the following decade, the hindcast period. For this, we rank ensemble members by their performances during the initialization and hindcast periods (lower rank is better and indicates less difference between observed and simulated trend). We then use linear regression to test for significant relationships between the ranks of the initialization and hindcast periods. Finally, we calculate mean trends and their standard deviations separately for each GCM before calculating the overall mean and standard deviation so that systematic errors of GCMs with many ensemble members are not dominating the overall results. Thus, first the trends of each ensemble member of each model are averaged, and next the model mean trends are averaged into a multi-model mean trend, similar to the MME calculation described in van Oldenborgh et al. (2012). Eq. 2 defines the overall standard deviation, which is adjusted to accommodate the different means μ and standard deviations σ for each model j with $\overline{\mu(T)}$ being the multi-model mean of all trends T.

$$Std.Dev. = \sqrt{\left(\left(\sum_{j}^{N}\left(\sigma_j(T)\right) + \sum_{j}^{N}\left(\mu_j(T) - \overline{\mu(T)}\right)^2\right)/N\right)} \qquad (2)$$

3 Analysis of SLP Trends

3.1 SLP Trends in Initialization Periods

Table 2 shows area averaged 22-year SLP trends during DJF from 1979 to 2000 for HadSLP2 observations and the CMIP3 versions of the GISS-ER and GISS-EH models, which provide ensemble members with trends close to the observed (see Table A1 in Appendix 1). Also listed are the average trends and their standard deviations for all eight GISS ensemble members (also known as runs, abbreviated as r*n* in this study, with *n* referring to the ensemble number), and the average trends and their standard deviations for all GISS ensemble members except the two worst performing members (one if only three members exist). This step is motivated by the attempt to reduce the ensemble prediction error when members with poor performance during the initialization period are excluded from the MME prediction. The bottom two rows in Table 2 show results for all 38 ensemble members from the nine CMIP3 GCMs, and for all except 14 poorly performing ensemble members. A complete list of the results of all 38 ensemble members is shown in Table A1. From the nine available CCSM3 ensemble members, only seven are used, since r4 and r8 do not continue with A1B projections. The remaining CCSM3 ensemble members are thus ranked from 1 to 7.

The ensemble members from each model are sorted by average rank over all the five zones that are listed in Table 2 with each zone being equally weighted for the overall average, despite the higher latitude zones being smaller in area than lower latitude zones. Here it is suggested that higher latitude zones have a stronger signal in decadal SLP variability, therefore they receive stronger weights per area. If two ensemble members of the same GCM have the same average rank, such as r2 and r3 in GISS-ER, which both

Table 2: Area averaged SLP trends from 1979 to 2000 for HadSLP2 observations and the two best performing CMIP3 GCMs. Numbers in parenthesis indicate the rank of each ensemble member within each GCM. The total ranks also include the average rank value over the five zones. The last four rows show average and standard deviation for the GISS and CMIP3 MMEs using all ensemble members and all except the lowest performing ensemble members, respectively.

SLP (Pa yr^{-1})	80°S-50°S	50°S-20°S	20°S-20°N	20°N-50°N	50°N-80°N
HadSLP2	-4.91	1.31	-0.26	1.74	-6.59
GISS-ER r4 (1, 2.6)	-5.18 (1)	1.09 (1)	0.18 (1)	-1.36 (5)	4.00 (5)
GISS-ER r5 (2, 2.8)	-4.04 (2)	2.30 (3)	2.43 (5)	1.02 (1)	-11.24 (3)
GISS-ER r1 (3, 2.8)	-2.41 (5)	1.65 (2)	0.56 (2)	-0.04 (3)	-2.36 (2)
GISS-ER r3 (4, 3.4)	-6.94 (4)	2.58 (5)	0.60 (3)	3.53 (4)	-6.71 (1)
GISS-ER r2 (5, 3.4)	-3.12 (3)	0.32 (4)	-1.06 (4)	0.97 (2)	1.20 (4)
GISS-EH r1 (1, 1.6)	-5.13 (1)	3.14 (2)	0.16 (2)	1.07 (1)	-3.06 (2)
GISS-EH r3 (2, 2.2)	-1.47 (3)	-1.27 (3)	0.16 (1)	3.72 (3)	-6.27 (1)
GISS-EH r2 (3, 2.2)	-3.81 (2)	2.34 (1)	0.82 (3)	3.52 (2)	-10.81 (3)
Avg. (St. Dev.) 8 GISS	-3.90 (2.40)	1.50 (1.87)	0.46 (0.95)	1.80 (1.81)	-4.87 (5.39)
Avg. (St. Dev.) 5 GISS	-3.59 (2.50)	1.31 (2.34)	0.61 (0.98)	1.13 (1.81)	-3.93 (5.78)
Avg. (St. Dev.) 38 CMIP3	-6.15 (8.19)	1.37 (2.99)	0.35 (0.88)	0.90 (2.65)	-0.39 (6.10)
Avg. (St. Dev.) 24 CMIP3	-6.85 (5.92)	1.84 (2.19)	0.43 (0.92)	1.09 (2.58)	-1.18 (5.98)

have an average rank of 3.4 (see Table 2), then the ensemble member with the larger spread of ranks over the five zones is considered to be favorable for predictions. In this case, r3 is ranked better because of the wider spread from 1 to 5 compared to r2 with a narrower spread from 2 to 4. A wider spread is considered to be favorable for predictions, because it increases the chance that the spread interval includes trends close to observations for some zones. If two ensembles have the same spread, a conservative approach is taken that assumes the ranks between initialization and hindcast do not match. In Tables 1 and A1 to A8, the different models are sorted with the best model during initialization appearing first. For this rank, the lowest detected SLP errors for each GCM and each zone (marked red in Tables A1, A3, A5, and A7) are averaged over the five zones.

After excluding the lowest performing ensemble members from the GISS-only MME, the average trend for 1979-2000 improves over the Southern Hemisphere midlatitudes, but slightly deteriorates over all other four zones. For the MME for all nine GCMs, the average trend for this period improves over the Northern Hemisphere but deteriorates over the Southern Hemisphere and the Tropics. In general, all averages remain within one standard deviation of the observations, apart from the Arctic zone, where the MME for all nine GCMs does not capture the strong negative trend. However, excluding poorly performing ensemble members decreases this error so that the strong trend over the Arctic zone is included within one standard deviation of the MME trend. As expected for fewer ensemble members, the standard deviation decreases in four out of five zones for the MME of all nine GCMs, which in the present case reduces the MME uncertainty. Based on these results with commonly studied GCMs, one might assume that both MME

error and standard deviation can be decreased when poor performing ensemble members are excluded for decadal predictions.

To further illustrate the selection process during the 1979-2000 period, the observed trends (Fig. 2a) are compared to the excluded ensemble members of GFDL-CM2.1 (Fig. 2b) and ECHAM5/MPI-OM (Figs. 2c-d). Black dots in Fig. 2 indicate significant trends at the 95% confidence interval based on Santer et al. (2000). The ensemble members in Fig. 2 fail to simulate the strength of the negative trend in the northern high-latitudes and only ECHAM5/MPI-OM r4 captures the slight positive trend in the northern mid-latitudes (see also Table A1). None of these ensemble members produce an increase in SLP over the southern mid-latitudes, and GFDL-CM2.1 r1 and ECHAM5/MPI-OM r4 also do not capture the simultaneous decrease over southern high-latitudes indicating the strengthening of SAM during this time period. The trend over the tropics is only slightly negative in HadSLP2 and stronger negative in these ensemble members, but SLP variability is relatively low in the tropics (Gillett and Stott 2009), and thus not discussed in detail here.

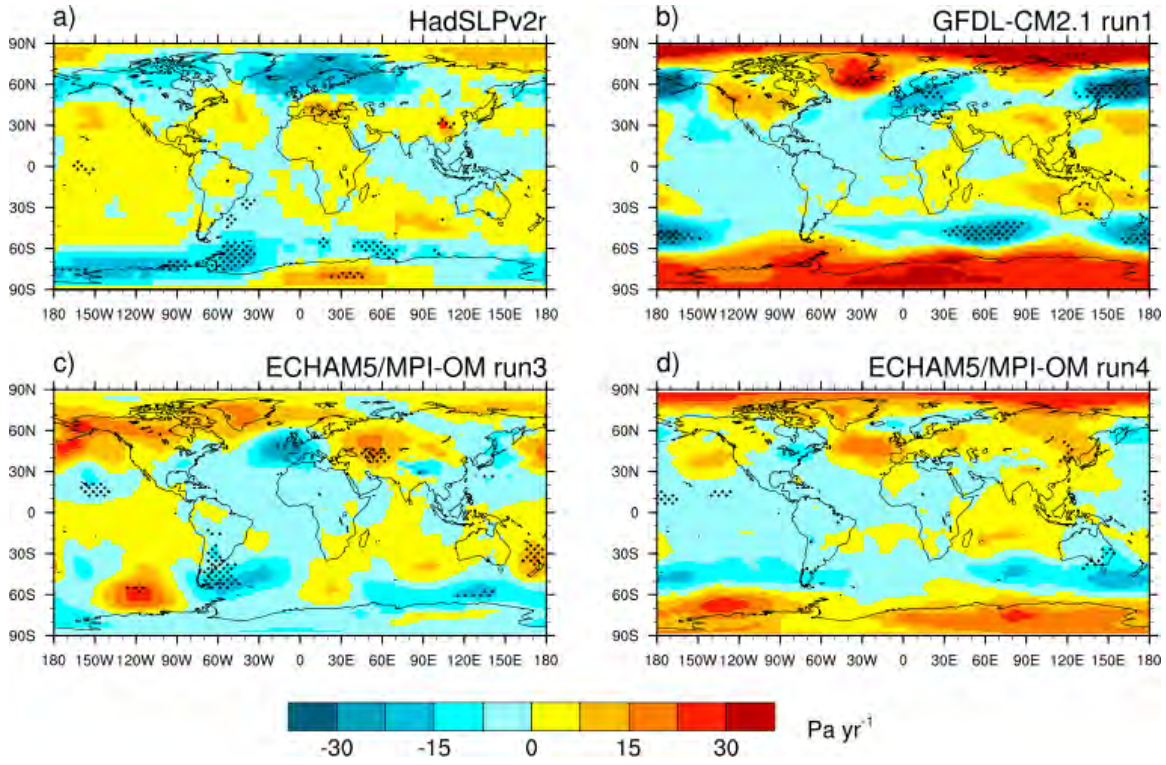

Fig. 2: SLP trends during DJF for 1979-2000 from **a** HadSLP2 observations and **b-d** individual GFDL-CM2.1 and ECHAM5/MPI-OM ensemble members that are found to have zonal trends different from HadSLP2. Black dots indicate significant trends at the 95% confidence interval.

The remaining ensemble members of these two models produce more realistic trend patterns (Fig. 3). For example, these ensemble members simulate the increased strength in SAM as shown by the negative (positive) trend over high- (mid-) latitudes. However, only GFDL-CM2.1 r2 simulates a negative trend in northern high-latitudes whereas other

ensemble members produce an erroneous positive trend, and all ECHAM5/MPI-OM ensemble members apart from r4, which has been excluded because of poor performance over the Southern Hemisphere, generate an erroneous negative trend in northern mid-latitudes. These erroneous trends explain why these two models are ranked low in Table 2, only followed by MIROC3.2(medres), which produces overly strong trends over the Southern Hemisphere (see Table A1).

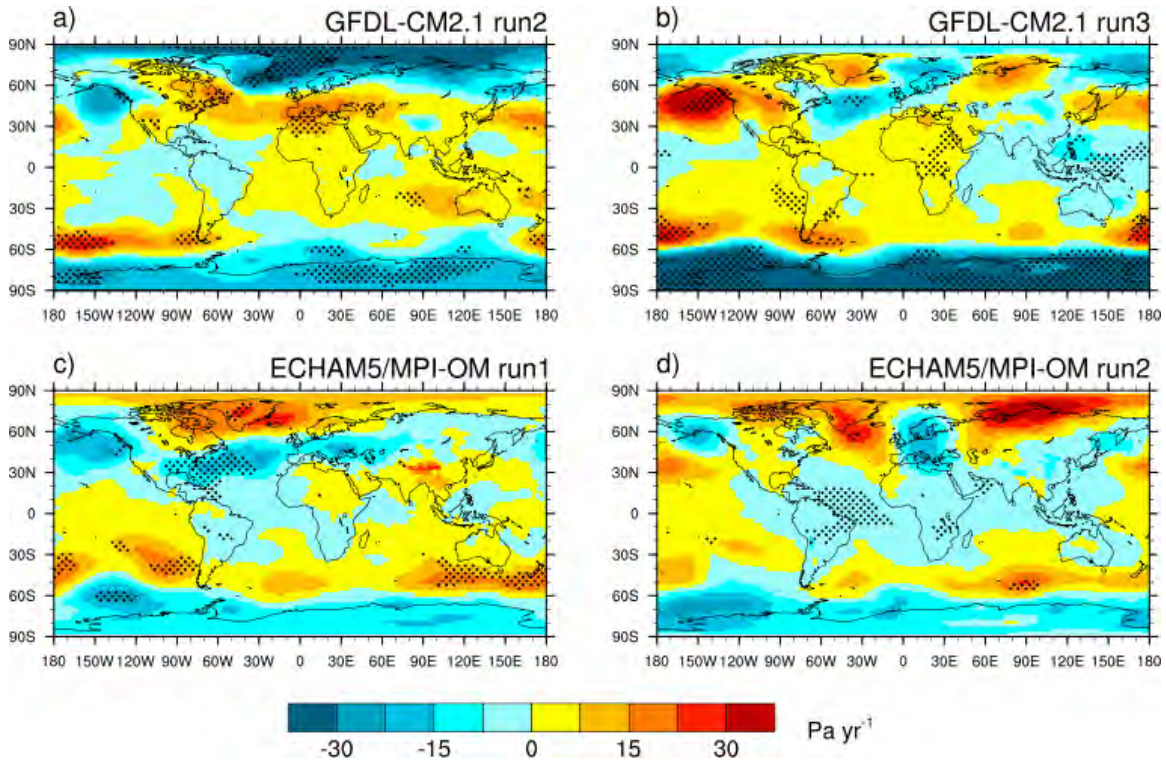

Fig. 3: As Fig. 2, but **a-b** for individual GFDL-CM2.1 and **c-d** ECHAM5/MPI-OM ensemble members that are found to have zonal trends similar to HadSLP2.

Decadal CMIP5 ensemble members are initialized with initial conditions reflecting the conditions of a particular year and generally cover a 30-year simulation period. The 1980 decadal experiment starts on January 1st 1981, and 1981-2000 is selected here as initialization period with the remaining ten years being used as hindcast period. The ranking of CCSM4 ensemble members results from the model's performance over the Northern Hemisphere, whereas CanCM4, MIROC5 and MPI-ESM-LR are mostly ranked related to their Southern Hemisphere performance. Ranking for the other three models (FGOALS-g2, MIROC4h, and MRI-CGCM3) results from performance over both hemispheres.

Table 3 shows HadSLP2 trends and average trends for all 38 ensemble members from all seven CMIP5 models and the best 20 ensemble members from all seven CMIP5 models. Trend anomalies were taken with respect to the 1981-2000 initialization period. Average trends were calculated first for each model and then averaged over the number of models. Average trends were improved over two of the five zones for all seven CMIP5 GCMs, indicating no significant improvements after excluding the worst performing ensemble members.

Table 3: Area averaged SLP trends from 1981 to 2000 for HadSLP2 trends, average trend anomalies for all CMIP5 ensemble members, and all but the lowest performing ensemble members. Standard deviations are in parenthesis.

SLP Trend (Pa yr^{-1})	80°S-50°S	50°S-20°S	20°S-20°N	20°N-50°N	50°N-80°N
HadSLP2 (1981-2000)	-5.05	0.88	0.01	0.48	-4.34
Avg. (St. Dev.) 38 CMIP5	0.30 (6.50)	0.30 (2.41)	0.20 (1.36)	-0.03 (3.94)	4.71 (10.60)
Avg. (St. Dev.) 28 CMIP5	1.03 (4.58)	0.19 (2.20)	0.37 (0.94)	0.27 (3.75)	3.39 (9.36)

3.2 SLP Trends in Hindcast Periods

In the HadSLP2 observations, the 11-year SLP trend for years 2001 to 2011 shows a much stronger negative signal over the high-latitude Southern Hemisphere and a slightly positive trend over the high-latitude Northern Hemisphere (Table 4), explaining the different characteristics of polar amplification over both hemispheres during DJF (see also Simmonds 2015). For CMIP3, the A1B scenario of each GCM is depicted in Table A2.

Table 4: Area averaged SLP trends as the summary in the last four lines of Table 2 compared to HadSLP2 observations, but for the hindcast period 2001-2011.

SLP (Pa yr^{-1})	80°S-50°S	50°S-20°S	20°S-20°N	20°N-50°N	50°N-80°N
HadSLP2	-26.83	7.23	-2.05	8.01	7.91
Avg. (St. Dev.) 8 GISS	-0.95 (13.82)	-1.90 (5.74)	-0.37 (2.55)	5.24 (4.55)	-6.03 (8.80)
Avg. (St. Dev.) 5 GISS	-6.17 (9.35)	0.56 (2.99)	0.06 (1.86)	5.10 (3.99)	-4.85 (9.08)
Avg. (St. Dev.) 38 CMIP3	-6.02 (15.22)	-0.97 (7.05)	0.91 (2.45)	2.93 (10.35)	-2.23 (14.23)
Avg. (St. Dev.) 24 CMIP3	-6.89 (15.48)	-0.34 (8.24)	1.07 (2.53)	0.68 (10.86)	0.86 (15.28)

It should be noted that the GFDL-CM2.1 A1B ensemble members downloadable from the data repository are inconsistent with the 20th century ensemble members: r1 continues as r3, r2 continues as r1, and r3 continues as r2. This study keeps the 20th century labeling for the A1B scenarios in Table A2 for purposes of consistency. The other ensemble members remain labeled as in the data repository.

For both GISS models, FGOALS-g1.0, GFDL-CM2.1 and MIROC3.2(medres), the worst performing ensemble members between years 1979 and 2000 maintain their ranks in the A1B prediction, and for CCSM3 and ECHAM5/MPI-OM, the single worst performing ensemble member maintains its rank in the A1B prediction. Only CGCM3.1(T47) and MRI-CGCM2.3.2 do not show any consistency between the two

scenarios. However, both excluded CGCM3.1(T47) ensemble members and the ECHAM5/MPI-OM r4 perform well over the northern mid-latitudes during the 1979-2000 initialization period (see Table A1) and maintain this performance during the 2001-2011 period.

In total, about half of the CMIP3 ensemble members (17 out of 38) maintain their ranks when grouped by GCMs (see Table A2). Furthermore, 65 out of 190 individual trends maintain their rank and 33 of the trends for models with more than three ensemble members shifted only by one rank. However, this pattern is less prevalent in decadal CMIP5 simulations for 1981-2010 (1961-1990) with only 10 (4) out of 38 ensemble members and 41 (18) out of 190 individual ensembles maintaining their rank in the decadal simulations (see Tables A4 and A8). Also, 24 (25) of the trends for models with more than three ensemble members shifted by only one rank.

We perform a regression analysis for the ranks during the initialization versus hindcast periods (Fig. 4) by normalizing the ranks so that the lowest rank for each model is equal to 7 (CMIP3) and 10 (CMIP5), respectively. The remaining ranks are spaced evenly between 1 and the lowest rank, thus allowing non-integer ranks in Figure 4. Because results from different GCMs, or samples, are combined, the total number of degrees of freedom for this significance test is reduced to $(n_1 + n_2 + ... + n_k) - k$, where n is the number of ensemble members for each sample and k is the number of samples. The regression analysis only shows a significant relationship between the initialization and the hindcast period for the 1979-2000 CMIP3 ensembles (Fig. 4a), not for the CMIP5 ensembles (Fig. 4b).

We address the discrepancy between the CMIP3 and CMIP5 simulations during the satellite era by additionally analyzing results from the same GCMs but for time periods just before the satellite era, namely CMIP3 data from 1957-1989 (Fig. 4c) and decadal CMIP5 data from 1961-1990 (Fig. 4d). The observed AMO was mostly in the negative phase during this earlier period. In both CMIP versions, we have found no significant relationship between these earlier initialization and hindcast periods, which leads to the conclusion that the significant relationship between the initialization and hindcast period that is found in CMIP3 models between 1979 and 2011 is a spurious result. It should be noted that even for the 1979-2000 CMIP3 ensembles none of the individual regions show a significant relationship (see Fig. A4). Although the CMIP5 results show significant trends over the northern midlatitudes and the tropical region (see Figs. A5d and A5e), it should also be noted that these results can occur randomly, such as the spurious significance of the negative relationship in the CMIP3 ensembles over the northern high-latitudes during 1957-1989 (Fig. A6f).

In addition to the spurious result that ranks are maintained in CMIP3 during the 1979-2011 period, the average values and standard deviations do not change significantly in these hindcast experiments after excluding low performing ensemble members. Thus, ranking ensemble members by their past performance may not be a suitable method to improve the climate prediction skill. Similarly, selecting only one or a subset of ensemble members per model may underrepresent the actually simulated climate variability of that model.

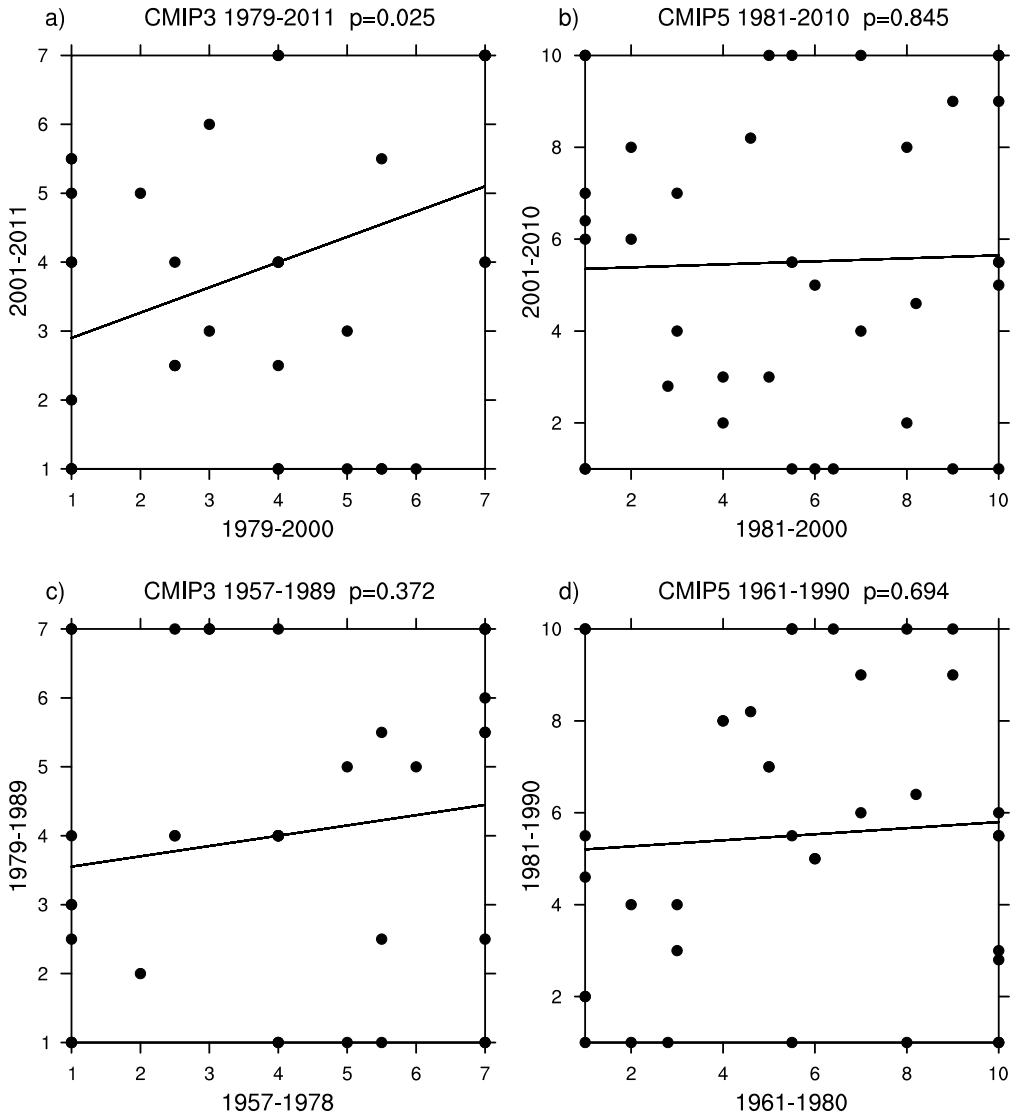

Fig. 4: Regression between the ensemble ranks in DJF initialization and hindcast periods for **a** CMIP3 during 1979-2011, **b** CMIP5 during 1981-2010, **c** CMIP3 during 1957-1989, and **d** CMIP5 during 1961-1980. *p* values are based on reduced degrees of freedom from multiple GCMs, or samples.

Different GCMs produce not only a different mean climate, but also a different variability between ensemble members. Fig. 5 shows SLP trends for the 2001 – 2011 period. When comparing HadSLP2 observations (Fig. 5a) to the CMIP3 multimodel mean trend (Fig. 5b) that includes all available CMIP3 simulations (van Oldenborgh 2016), it becomes obvious that the trends in the multimodel mean are up to one order of magnitude lower than observed. Applying the exclusion process to the nine selected models described in Section 2 (Fig. 5c) suggests some minor improvement compared to including all ensemble members (Fig. 5d), such as the positive trend around Greenland and the reduced erroneous positive trend over the Sargasso Sea in the North Atlantic, but the negative trend over western Europe and the positive trend over the Middle East are

not as well represented. In general, the nine-model ensemble shows a similar spatial variability to the CMIP3 multimodel mean for large-scale trend detections. For the two models with large similarity between initialization and hindcast rankings, GISS-ER and GISS-EH, the spatial variability remains similar between all and all but excluded ensemble members (Figs. 5e-f). However, it should be noted that these models were ranked low in a model comparison study with regard to their climate representation (Gleckler et al. 2008). On the other hand, the example of GFDL-CM2.1 and ECHAM5/MPI-OM (Figs. 5g-h) shows a strong increase in spatial variability of decadal prediction after excluding poor performing ensemble members: the general strength of trends is improved and trends over many regions such as the west coast of the Americas, the Middle East, central Russia, and Australia appear closer to the observed. However, other regions such as Western Europe remain in the wrong phase, making the interpretation of these trends difficult due to the large spatial inaccuracies in many regions. Also, trends in many subsets of the MME are largely non-significant.

Because individual ensemble members can have different probabilities for the likelihood that the trend is significant, we used Fisher's combined probability test (Mosteller and Fisher 1948) to identify the likelihood that the previously described SLP trends are statistically significant in an ensemble average:

$$\chi^2_{2k} = -2\sum_{i=1}^{k} \ln(p_i) \quad (3)$$

with p_i being the p-values for trends of each ensemble member i, with k being the number of ensemble members, and with χ^2_{2k} being the χ^2 value for $2k$.

Predictions with decadal CMIP5 ensemble members (Fig. 6) are compared to a MME that includes the seven CMIP5 models listed in Table A3 (Fig. 6c-d) and the MME of all available historical CMIP5 simulations (van Oldenborgh 2016) (Fig. 6b). Compared to the CMIP3 multimodel mean SLP trends (Fig. 5b), significant negative trends over South America and around southwest Africa and the strength of positive trends over the North and South Pacific are better represented in the CMIP5 MME, and positive trends over Greenland and negative trends over Europe and eastern Antarctica are less well simulated by the historical experiments during this decade. Despite the large differences in HadSLP2 observations when adding a single year (2011) to the trend calculations (Figs. 5a and 6a) that indicate the strong year-to-year variability in seasonal SLP values, CMIP5 decadal predictions (Figs. 6c-f) appear to simulate improved trend patterns compared to the multi-model mean. CanCM4 and FGOALS-g2 show similar trends to observations (see Table A3) and FGOALS-g2 and CanESM2, which has a similar setup to CanCM4, simulate relatively realistic SAM patterns (Zheng et al. 2013). Boer et al. (2013) used CanCM4 to conclude that the effect of initialization diminishes beyond about a three-year average and the skill of the initialized forecast begins to approach that of the uninitialized simulation with both exhibiting modestly increasing skill with increasing averaging time. However, removing worse performing ensemble members does not affect the global SLP trend pattern in this subset of models (Fig. 6e-f), which corroborates the hypothesis that the significant relationships between initialization and hindcast periods in the CMIP3 patterns are spurious.

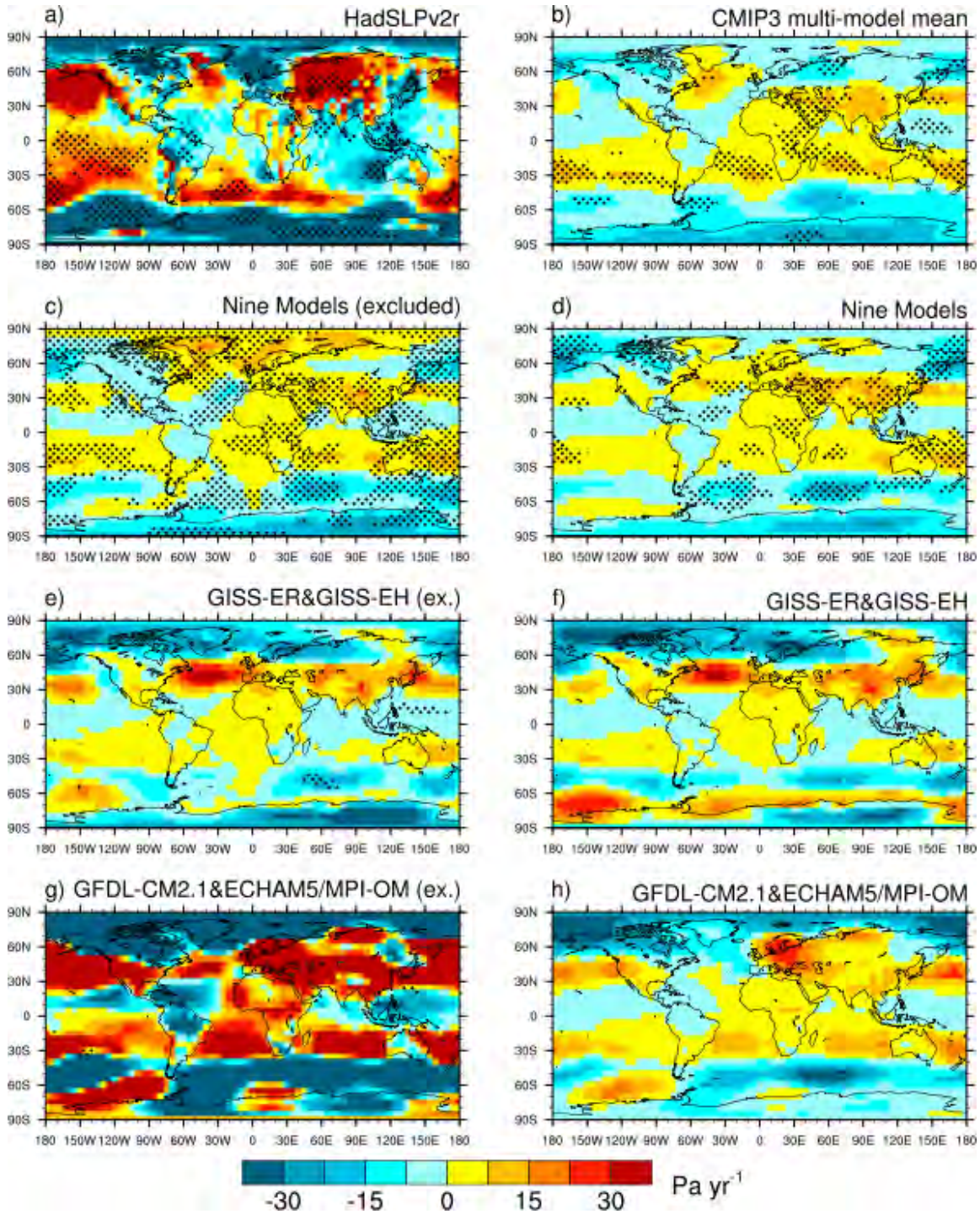

Fig. 5: SLP trends during DJF for 2001-2011 from **a** HadSLP2 observations, **b** the CMIP3 MME, **c** a subset of ensemble members from nine GCMs, **d** as **c** but all ensemble members from nine GCMs, **e** a subset of ensemble members from GISS-ER and GISS-EH, **f** as **e** but all ensemble members from GISS-ER and GISS-EH, **g** a subset of ensemble members from GFDL-CM2.1 and ECHAM5/MPI-OM, **h** as **g** but all ensemble members from GFDL-CM2.1 and ECHAM5/MPI-OM.

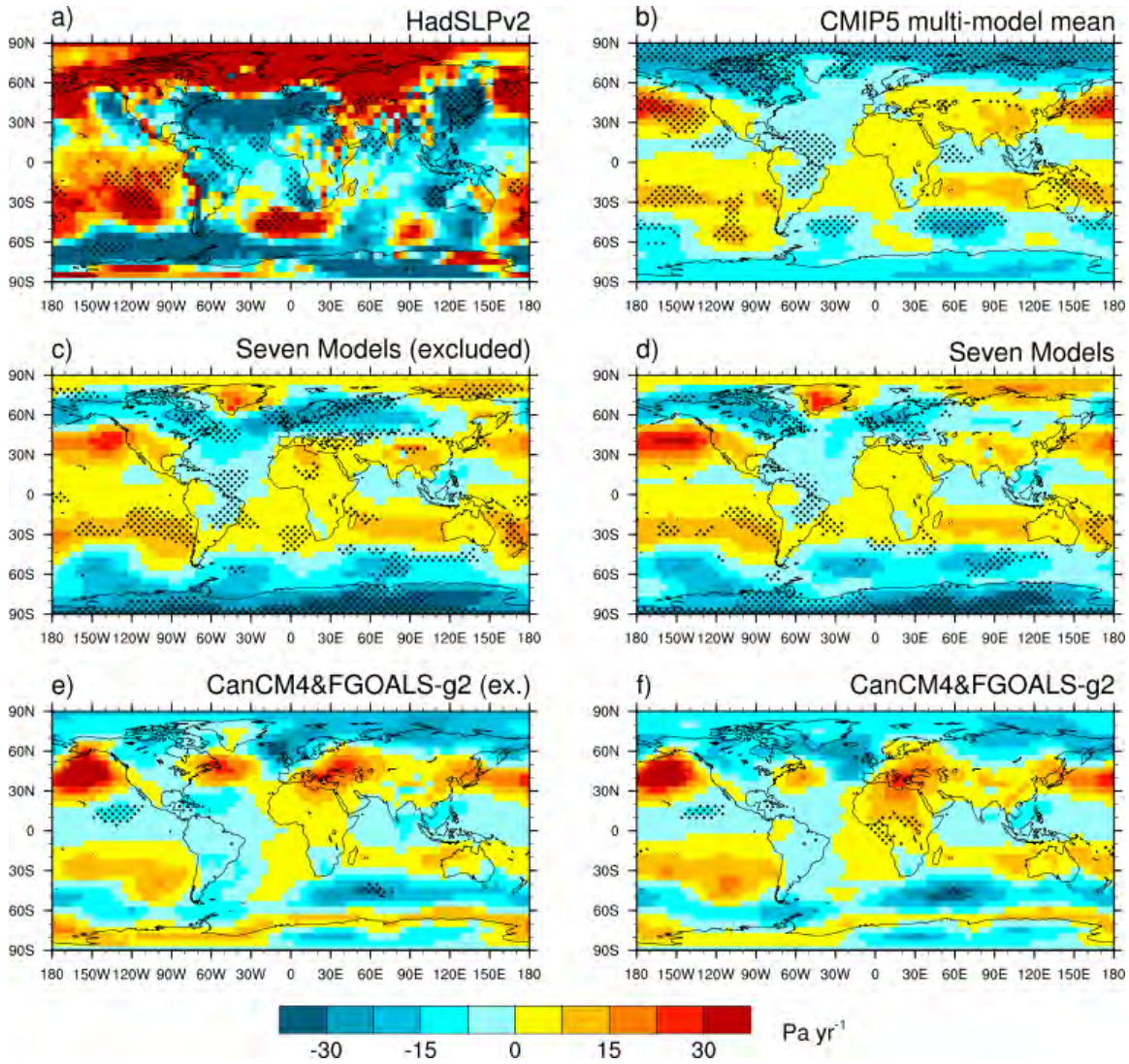

Fig. 6: As Fig. 5, but for the 2001-2010 trends of CMIP5 historical multi-model mean and CMIP5 decadal ensemble members.

Table 5 shows HadSLP2 trends and trend anomalies for the same ensemble members as in Table 3, but for the hindcast period. As in Fig. 6, trend anomalies were taken with respect to the 2001-2010 hindcast period. The CMIP5 MMEs are less sensitive to the selection of appropriate ensemble members and the standard deviations in CMIP5 are generally higher than in CMIP3. Similar to Table 3, Table 5 does not indicate an improvement in the average simulated trend after excluding the worst performing CMIP5 ensemble members. It is suggested that this increase in variability is related to the representation of more processes in greater detail (Knutti and Sedlacek 2013).

Table 5: As Table 3, but for the 10-year hindcast period 2001-2010.

SLP Trend (Pa yr^{-1})	80°S-50°S	50°S-20°S	20°S-20°N	20°N-50°N	50°N-80°N
HadSLP2 (2001-2010)	-16.61	4.57	-3.65	-10.36	34.81
Avg. (St. Dev.) 38 CMIP5	5.94 (16.38)	-2.86 (8.87)	3.70 (4.85)	13.09 (13.68)	-37.45 (29.34)
Avg. (St. Dev.) 28 CMIP5	6.59 (15.02)	-2.39 (7.21)	3.45 (4.74)	12.63 (14.01)	-37.70 (30.58)

4 Summary and Conclusion

In this study, we describe a spurious pattern in decadal SLP trends as simulated by a subset of CMIP3 models that is only present during the satellite era of 1979-2011. We find that such trends are not present in CMIP3 model simulations during the earlier 1957-1989 period or in the more recent decadal CMIP5 predictions. This warrants caution when interpreting decadal prediction skill based on limited data. Consequently, we also show that results from a hindcast period of ensemble prediction experiments do not necessarily benefit from excluding ensemble members with poor performance during an initialization period.

The performance of individual ensemble members is described by ranking them based on their representation of global patterns in SLP. Previous studies have shown the importance of realistic SST initializations for decadal predictions (Keenlyside et al. 2008). For example, van Oldenborgh et al. (2012) found in a comparison of a four-model prediction with the CMIP3 MME that the decadal skill in the northern North Atlantic and eastern Pacific is most likely due to model initialization, whereas the skill in the subtropical North Atlantic and western North Pacific are likely due to the GHG and aerosol forcing. Yeager et al. (2012) suggested that the trend of ocean heat content between the subpolar and subtropical gyres in the North Atlantic is strongly influential on large-scale climate variability. The method in the present paper tests if there is a relationship between SLP patterns in decadal model initializations and the following decade, or hindcast period. In the presented example, a statistically significant relationship is detected whereby half of the CMIP3 ensemble members retain their ranks based on a 22-year initialization or training period during a hindcast or testing period of 11 years. Although we were able to identify such a relationship for the CMIP3 1979-2011 period, this relationship cannot be verified during the earlier 1957-1989 period, where 14 out of 38 ensemble members retain their ranks (see Appendix 1). These numbers are greatly reduced in decadal CMIP5 simulations, indicating an increase in spatial SLP variability in CMIP5 compared to CMIP3. Thus, despite the statistically significant relationship in the 1979-2011 CMIP3 simulations, the predictive value of the present approach cannot be corroborated during other periods or with CMIP5 models.

The reduced relationship for the 1957-1989 period is likely a result of the opposite AMO phase and lower trends during the initialization period, where the shift from positive to negative AMO phase in the early 1960's (Sutton and Hodson 2005; Denton and Broecker 2008) is counteracting the warming over high-latitude regions due to GHG emissions (Hegerl et al. 1997; Dima and Lohmann 2007). This results in reduced trends

during the initialization period, and thus reduced skill during the hindcast period, when compared to the later 1979-2011 period.

In summary, this study highlights the importance of model intercomparison projects, not only between different GCMs, but also between different generations of models and different simulation time periods for model evaluation. Because of the significant influence of decadal climate variability on human well-being and our natural resources, it is important that predictions of future climate change accurately portray the plausible outcomes.

Acknowledgments

HadSLP2 data were received from the UK Met Office Hadley Centre Observations website. CMIP3 data were received from the PCMDI server and CMIP5 data were obtained from the Earth System Grid. Multi-model mean values for both CMIP3 and CMIP5 were retrieved from the KNMI Climate Explorer. The trend analysis was performed at the Minnesota Supercomputing Institute.

An earlier version of this manuscript benefited from the help of undergraduate researchers D. Ormsby and G. D. Smith at the University of Minnesota. Support for this study was provided by the U.S. National Science Foundation under grant 1029711, the U.S. National Aeronautics and Space Administration under Grant 14-CMAC14-0010, and the George R. and Orpha Gibson Foundation at the University of Minnesota.

Appendix 1

This appendix contains extensive tables and figures of the trend analysis in CMIP3 and CMIP5 ensemble members. Tables A1-A2 provide observed and simulated SLP trends for all analyzed CMIP3 ensemble members during DJF for the periods 1979 – 2000 and 2001 – 2011, which are described in the main manuscript (see Tables 2 and 4). 17 (3) out of 38 runs, and 62 (33) out of 190 regional estimates retain their ranks (neighboring ranks) from 1979 – 2000 in the 2001 – 2011 predictions, which leads to a statistically significant, but spurious relationship in global SLP pattern between the 1979 – 2000 and the 2001 – 2011 period. Tables A3 and A4 show results for the 30-year CMIP5 decadal runs starting in 1981 (see Tables 3 and 5), where only 10 (3) out of 38 runs, and 34 (24) out of 190 regional estimates retain their ranks (neighboring ranks) from 1981 – 2000 in the 2001 – 2010 predictions.

Tables A5-A10 and Figs. A1-A3 complement these findings with independent results with CMIP3 (CMIP5) GCMs for the training periods 1957 – 1978 (1961 – 1980) and the predictions for 1979 – 1989 (1981 – 1990) to verify the proposed hypothesis with independent time periods. In the CMIP3 case, 14 out of 38 models, and 47 (32) out of 190 regional estimates retain their ranks (neighboring ranks) from 1957 – 1978 in the 1979 – 1989 simulations. Variability decreases in most zones, and hindcast averages for the GISS models improve mainly over the Northern Hemisphere. However, GHG influence was less during the earlier period, and the AMO was mostly in a negative phase in the 1960's, resulting in a reduced trend and less global influence of the AMO. Detailed regression analyses in Figs. A4-A7 show how all four global trend analyses are largely based on non-significant trends over the five regions.

Table A1: CMIP3 trends for 1979-2000 (ensemble members used for ranking models are marked red, numbers in parenthesis indicate the rank of each ensemble member within each GCM).

SLP (Pa yr^{-1})	80°S-50°S	50°S-20°S	20°S-20°N	20°N-50°N	50°N-80°N
HadSLP2	-4.91	1.31	-0.26	1.74	-6.59
GISS-ER r4 (1)	-5.18 (1)	1.09 (1)	0.18 (1)	-1.36 (5)	4.00 (5)
GISS-ER r5 (2)	-4.04 (2)	2.30 (3)	2.43 (5)	1.02 (1)	-11.24 (3)
GISS-ER r1 (3)	-2.41 (5)	1.65 (2)	0.56 (2)	-0.04 (3)	-2.36 (2)
GISS-ER r3 (4)	-6.94 (4)	2.58 (5)	0.60 (3)	3.53 (4)	-6.71 (1)
GISS-ER r2 (5)	-3.12 (3)	0.32 (4)	-1.06 (4)	0.97 (2)	1.20 (4)
GISS-EH r1 (1)	-5.13 (1)	3.14 (2)	0.16 (2)	1.07 (1)	-3.06 (2)
GISS-EH r3 (2)	-1.47 (3)	-1.27 (3)	0.16 (1)	3.72 (3)	-6.27 (1)
GISS-EH r2 (3)	-3.81 (2)	2.34 (1)	0.82 (3)	3.52 (2)	-10.81 (3)
CGCM3.1(T47) r1 (1)	-10.95 (2)	1.19 (1)	0.21 (1)	-1.93 (5)	7.57 (5)
CGCM3.1(T47) r3 (2)	1.32 (3)	-0.36 (2)	0.13 (2)	-1.07 (4)	-3.15 (3)
CGCM3.1(T47) r2 (3)	-5.58 (1)	-1.38 (5)	1.11 (4)	2.81 (3)	-4.12 (2)
CGCM3.1(T47) r4 (4)	-13.16 (4)	3.43 (4)	1.88 (5)	2.51 (2)	-4.54 (1)
CGCM3.1(T47) r5 (5)	-13.36 (5)	3.21 (3)	0.94 (3)	1.58 (1)	-1.13 (4)
CCSM3 r9 (1)	-7.38 (2)	0.74 (2)	-0.11 (2)	2.64 (2)	1.31 (4)
CCSM3 r6 (2)	-7.66 (3)	1.43 (1)	-0.26 (3)	2.89 (4)	2.16 (5)
CCSM3 r5 (3)	-6.31 (1)	-0.28 (4)	-1.43 (7)	2.64 (1)	6.42 (6)
CCSM3 r3 (4)	-9.71 (4)	3.67 (6)	0.81 (4)	5.10 (7)	-9.64 (1)
CCSM3 r7 (5)	-14.98 (7)	3.13 (5)	0.29 (1)	4.30 (6)	0.94 (3)
CCSM3 r1 (6)	-11.46 (5)	4.55 (7)	0.88 (5)	0.13 (5)	-1.21 (2)
CCSM3 r2 (7)	-14.40 (6)	2.10 (3)	-0.89 (6)	2.74 (3)	8.67 (7)
MRI-CGCM2.3.2 r1 (1)	-8.29 (1)	1.43 (1)	0.75 (3)	1.47 (2)	0.86 (4)
MRI-CGCM2.3.2 r2 (2)	-0.88 (3)	-0.17 (2)	-0.30 (4)	2.13 (3)	-4.60 (1)
MRI-CGCM2.3.2 r4 (3)	0.68 (4)	-1.13 (3)	0.56 (2)	-0.23 (4)	-1.17 (2)
MRI-CGCM2.3.2 r5 (4)	-1.28 (2)	-1.60 (5)	0.40 (1)	-0.49 (5)	2.31 (5)
MRI-CGCM2.3.2 r3 (5)	-15.49 (5)	4.16 (4)	1.03 (5)	1.99 (1)	0.83 (3)
FGOALS-g1.0 r3 (1)	-6.85 (1)	1.72 (1)	0.83 (1)	1.23 (1)	-3.79 (1)
FGOALS-g1.0 r1 (2)	-11.49 (2)	1.87 (2)	1.18 (3)	-1.10 (2)	3.90 (2)
FGOALS-g1.0 r2 (3)	7.96 (3)	-3.72 (3)	-0.42 (2)	-3.19 (3)	4.66 (3)
GFDL-CM2.1 r2 (1)	-5.81 (1)	2.24 (1)	1.11 (3)	4.40 (2)	-9.19 (1)
GFDL-CM2.1 r3 (2)	-15.22 (2)	4.65 (2)	-0.06 (1)	3.37 (1)	1.00 (3)
GFDL-CM2.1 r1 (3)	9.97 (3)	-3.72 (3)	-0.37 (2)	-1.10 (3)	-0.75 (2)
ECHAM5/MPI-OM r1 (1)	-5.08 (1)	5.31 (3)	0.33 (1)	-4.40 (4)	0.22 (1)
ECHAM5/MPI-OM r2 (2)	-6.37 (2)	2.25 (1)	-1.36 (4)	-0.36 (2)	4.51 (3)
ECHAM5/MPI-OM r4 (3)	4.50 (4)	-5.06 (4)	-0.89 (3)	1.31 (1)	3.75 (2)
ECHAM5/MPI-OM r3 (4)	-3.31 (3)	-0.07 (2)	-0.41 (2)	-1.13 (3)	6.23 (4)
MIROC3.2(medres) r2 (1)	-9.45 (1)	3.59 (2)	0.62 (1)	2.60 (1)	-3.94 (1)
MIROC3.2(medres) r3 (2)	-14.37 (2)	2.99 (1)	1.21 (3)	0.26 (2)	4.35 (2)
MIROC3.2(medres) r1 (3)	-18.67 (3)	4.81 (3)	0.69 (2)	-3.45 (3)	11.97 (3)

Table A2: CMIP3 trends for 2001 – 2011 (ensemble members that retain their ranks are marked green, ensembles with a neighboring rank are marked yellow).

SLP (Pa yr^{-1})	80°S-50°S	50°S-20°S	20°S-20°N	20°N-50°N	50°N-80°N
HadSLP2	-26.83	7.23	-2.05	8.01	7.91
GISS-ER r4 (1)	-3.81 (3)	4.32 (1)	1.34 (3)	7.04 (1)	-19.51 (5)
GISS-ER r5 (3)	-2.72 (4)	-2.04 (4)	-2.54 (1)	7.04 (2)	-4.03 (3)
GISS-ER r1 (2)	-7.18 (1)	1.34 (2)	2.36 (4)	-0.96 (4)	-0.12 (2)
GISS-ER r3 (4)	-0.84 (5)	0.75 (3)	-0.26 (2)	-1.66 (5)	2.51 (1)
GISS-ER r2 (5)	-4.70 (2)	-2.66 (5)	3.53 (5)	1.62 (3)	-9.28 (4)
GISS-EH r1 (2)	1.21 (2)	-1.96 (2)	-0.07 (2)	4.21 (2)	-6.00 (2)
GISS-EH r3 (1)	-16.77 (1)	1.79 (1)	-0.46 (1)	7.43 (1)	2.37 (1)
GISS-EH r2 (3)	21.40 (3)	-12.27 (3)	-4.37 (3)	11.94 (3)	-14.31 (3)
CGCM3.1(T47) r1 (4)	2.87 (3)	1.17 (1)	3.64 (5)	-8.91 (5)	-6.90 (5)
CGCM3.1(T47) r3 (2)	-2.17 (2)	0.14 (2)	-0.32 (3)	-6.87 (4)	6.69 (1)
CGCM3.1(T47) r2 (5)	25.66 (5)	-13.60 (5)	-0.11 (4)	-5.37 (3)	5.38 (2)
CGCM3.1(T47) r4 (1)	-4.00 (1)	-11.81 (4)	-1.156 (2)	15.00 (1)	-5.02 (4)
CGCM3.1(T47) r5 (3)	12.89 (4)	-4.77 (3)	-1.56 (1)	0.67 (2)	-4.18 (3)
CCSM3 r9 (2)	-10.04 (3)	-4.00 (5)	0.20 (2)	11.41 (3)	-2.48 (3)
CCSM3 r6 (5)	-4.87 (5)	6.30 (1)	0.99 (3)	0.81 (5)	-13.48 (5)
CCSM3 r5 (6)	-6.36 (4)	5.16 (3)	3.27 (6)	-13.60 (7)	7.12 (2)
CCSM3 r3 (4)	-14.63 (1)	10.71 (4)	2.10 (4)	0.84 (4)	-15.76 (6)
CCSM3 r7 (3)	9.16 (7)	-10.17 (7)	-3.91 (1)	4.89 (2)	7.25 (1)
CCSM3 r1 (1)	-13.92 (2)	6.22 (2)	2.44 (5)	10.23 (1)	-24.19 (7)
CCSM3 r2 (7)	0.88 (6)	-2.29 (5)	4.23 (7)	-2.14 (6)	-5.96 (4)
MRI-CGCM2.3.2 r1 (4)	-2.64 (4)	-0.74 (4)	3.45 (5)	-11.97 (4)	13.91 (2)
MRI-CGCM2.3.2 r2 (2)	-28.03 (1)	2.78 (2)	0.33 (3)	13.15 (1)	-8.75 (4)
MRI-CGCM2.3.2 r4 (5)	1.22 (5)	1.81 (3)	1.75 (4)	-19.39 (5)	23.65 (3)
MRI-CGCM2.3.2 r5 (1)	-19.78 (2)	-4.75 (5)	-0.78 (1)	13.30 (2)	5.55 (1)
MRI-CGCM2.3.2 r3 (3)	-35.12 (3)	6.81 (1)	-0.33 (2)	19.78 (3)	-13.23 (5)
FGOALS-g1.0 r3 (1)	-22.14 (1)	3.10 (1)	1.23 (2)	3.51 (1)	1.98 (1)
FGOALS-g1.0 r1 (2)	12.34 (3)	-6.45 (3)	-0.08 (1)	-15.22 (3)	23.84 (2)
FGOALS-g1.0 r2 (3)	-5.56 (2)	-0.50 (2)	1.48 (3)	13.91 (2)	-22.84 (3)
GFDL-CM2.1 r2 (2)	-18.95 (1)	7.26 (1)	4.25 (3)	11.61 (2)	-26.17 (3)
GFDL-CM2.1 r3 (1)	-2.05 (2)	-13.41 (3)	-1.54 (1)	10.04 (1)	10.17 (1)
GFDL-CM2.1 r1 (3)	6.69 (3)	-0.27 (2)	3.11 (2)	-0.74 (3)	-12.29 (2)
ECHAM5/MPI-OM r1 (3)	-8.92 (2)	5.20 (1)	2.01 (4)	-0.40 (4)	-6.18 (4)
ECHAM5/MPI-OM r2 (2)	-3.26 (3)	0.58 (4)	-0.85 (1)	2.16 (2)	1.12 (2)
ECHAM5/MPI-OM r4 (1)	-25.75 (1)	4.04 (2)	-0.49 (2)	7.50 (1)	9.93 (1)
ECHAM5/MPI-OM r3 (4)	2.49 (4)	0.63 (3)	0.69 (3)	-0.15 (3)	-5.97 (3)
MIROC3.2(medres) r2 (1)	-31.50 (1)	7.32 (1)	5.04 (3)	-7.94 (3)	16.49 (1)
MIROC3.2(medres) r3 (2)	-12.06 (2)	-9.56 (3)	2.146 (1)	12.67 (1)	-2.12 (2)
MIROC3.2(medres) r1 (3)	-10.42 (3)	-1.50 (2)	3.29 (2)	2.18 (2)	-2.18 (3)

Table A3: As Table A1, but for CMIP5 decadal simulations (1981-2000).

SLP (Pa yr^{-1})	80°S-50°S	50°S-20°S	20°S-20°N	20°N-50°N	50°N-80°N
HadSLP2	-5.05	0.88	0.01	0.48	-4.34
CCSM4 r5 (1)	-7.36 (5)	1.69 (3)	0.67 (5)	4.27 (6)	-4.52 (1)
CCSM4 r1 (2)	1.92 (10)	-0.03 (5)	0.18 (2)	-0.16 (2)	-2.35 (3)
CCSM4 r8 (3)	-1.94 (8)	0.8 (1)	-1.21 (7)	1.09 (1)	0.94 (6)
CCSM4 r3 (4)	-2.49 (7)	3.69 (7)	0.15 (1)	-2.06 (4)	-0.08 (5)
CCSM4 r7 (5)	-2.58 (6)	1.73 (4)	-0.17 (3)	5.82 (8)	-7.39 (4)
CCSM4 r10 (6)	-6.09 (2)	4.32 (9)	2.44 (9)	-2.12 (5)	-4.83 (2)
CCSM4 r9 (7)	-3.73 (3)	-0.74 (6)	-0.19 (4)	11.07 (10)	-13.27 (7)
CCSM4 r6 (8)	-5.9 (1)	4.98 (10)	2.49 (10)	2 (3)	-13.43 (8)
CCSM4 r2 (9)	-3.1 (4)	-2 (8)	-1.1 (6)	-4.06 (7)	19.21 (10)
CCSM4 r4 (10)	-1.63 (9)	1.4 (2)	1.49 (8)	8.01 (9)	-15.94 (9)
CanCM4 r4 (1)	-6.79 (3)	0.64 (1)	2.10 (10)	-1.42 (2)	-0.91 (2)
CanCM4 r5 (2)	-0.97 (6)	-1.22 (5)	0.56 (3)	-1.79 (3)	1.78 (5)
CanCM4 r10 (3)	-11.13 (8)	-0.08 (3)	0.73 (4)	6.17 (9)	-3.21 (1)
CanCM4 r3 (4)	-11.17 (9)	4.50 (9)	-0.13 (1)	-0.93 (1)	2.42 (6)
CanCM4 r7 (5)	-6.78 (2)	-1.44 (6)	0.38 (2)	-2.31 (6)	9.87 (10)
CanCM4 r2 (6)	0.32 (7)	-0.03 (2)	1.51 (8)	-4.62 (8)	-0.54 (3)
CanCM4 r9 (7)	-3.71 (1)	2.30 (4)	1.41 (7)	-5.64 (10)	4.18 (8)
CanCM4 r8 (8)	-2.23 (5)	-1.54 (8)	1.30 (6)	3.90 (7)	-8.61 (4)
CanCM4 r6 (9)	-3.05 (4)	-3.12 (10)	-0.91 (5)	3.26 (5)	5.07 (9)
CanCM4 r1 (10)	1.85 (10)	-1.49 (7)	1.91 (9)	2.87 (4)	-11.54 (7)
FGOALS-g2 r2 (1)	-5.28 (1)	-0.21 (1)	-0.73 (1)	-0.80 (1)	10.85 (3)
FGOALS-g2 r3 (2)	-4.41 (2)	2.17 (2)	0.97 (2)	3.95 (3)	-9.68 (2)
FGOALS-g2 r1 (3)	7.30 (3)	-2.73 (3)	-2.09 (3)	3.76 (2)	-3.26 (1)
MIROC4h r2 (1)	-0.78 (1)	0.69 (1)	0.58 (1)	0.90 (1)	-5.91 (1)
MIROC4h r1 (2)	-0.70 (2)	-2.75 (3)	0.86 (2)	2.21 (2)	-2.58 (2)
MIROC4h r3 (3)	-10.87 (3)	2.10 (2)	2.63 (3)	4.47 (3)	-10.66 (3)
MIROC5 r2 (1)	-0.62 (4)	0.58 (2)	-0.15 (2)	-0.96 (4)	-0.38 (2)
MIROC5 r1 (2)	-5.17 (1)	1.08 (1)	1.04 (4)	-2.27 (5)	2.31 (5)
MIROC5 r6 (3)	-6.40 (2)	1.58 (3)	0.92 (3)	5.00 (6)	-8.62 (3)
MIROC5 r3 (4)	-7.65 (3)	3.25 (5)	1.47 (6)	1.86 (3)	-7.99 (1)
MIROC5 r5 (5)	-16.8 (6)	5.13 (6)	-0.07 (1)	1.04 (1)	4.07 (6)
MIROC5 r4 (6)	-0.10 (5)	2.18 (4)	1.31 (5)	-0.42 (2)	-8.67 (4)
MPI-ESM-LR r2 (1)	-3.39 (1)	2.97 (1)	-0.80 (3)	3.64 (3)	-6.64 (1)
MPI-ESM-LR r3 (2)	-13.6 (2)	3.56 (2)	0.05 (1)	0.05 (2)	4.59 (3)
MPI-ESM-LR r1 (3)	-14.12 (3)	5.00 (3)	-0.54 (2)	0.47 (1)	3.29 (2)
MRI-CGCM3 r3 (1)	-2.97 (2)	1.20 (1)	-0.59 (1)	-0.63 (1)	2.88 (1)
MRI-CGCM3 r1 (2)	-5.91 (1)	3.90 (3)	0.90 (2)	6.65 (3)	-14.11 (2)
MRI-CGCM3 r2 (3)	-14.35 (3)	2.05 (2)	-2.62 (3)	-5.48 (2)	24.52 (3)

Table A4: As Table A2, but for CMIP5 decadal simulations (2001-2010).

SLP (Pa yr^{-1})	80°S-50°S	50°S-20°S	20°S-20°N	20°N-50°N	50°N-80°N
HadSLP2	-16.61	4.57	-3.65	-10.36	34.81
CCSM4 r5 (6)	-18.08 (3)	4.14 (1)	4.38 (9)	8.93 (6)	-14.93 (7)
CCSM4 r1 (8)	-17.78 (2)	11.7 (7)	0.59 (4)	12.38 (8)	-29.39 (10)
CCSM4 r8 (7)	-7.88 (7)	7.74 (5)	0.96 (5)	1.74 (3)	-13.33 (6)
CCSM4 r3 (3)	-15.63 (1)	14.19 (9)	4.58 (10)	-14 (1)	-4.47 (4)
CCSM4 r7 (10)	-1.56 (8)	-7.35 (10)	0.51 (3)	16.77 (10)	-20.15 (8)
CCSM4 r10 (5)	8.26 (10)	2.23 (3)	1.04 (6)	-4.51 (2)	-7.09 (5)
CCSM4 r9 (4)	-8.85 (4)	-3.56 (8)	-6.67 (2)	16.33 (9)	3.78 (2)
CCSM4 r6 (2)	-8.47 (6)	1.25 (6)	-9.21 (7)	5.1 (4)	23.57 (1)
CCSM4 r2 (1)	-8.75 (5)	2.05 (4)	-2.41 (1)	9.96 (7)	-2.26 (3)
CCSM4 r4 (9)	0.65 (9)	2.67 (2)	3.32 (8)	7.08 (5)	-26.38 (9)
CanCM4 r4 (7)	11.11 (7)	-1.93 (4)	4.52 (8)	0.14 (3)	-28.02 (7)
CanCM4 r5 (6)	-28.29 (4)	14.77 (8)	2.78 (7)	0.36 (4)	-16.75 (6)
CanCM4 r10 (4)	1.02 (6)	1.11 (2)	-3.65 (1)	23.26 (10)	-30.73 (8)
CanCM4 r3 (2)	-2.44 (5)	-2.19 (5)	-3.00 (3)	4.98 (6)	-0.62 (1)
CanCM4 r7 (3)	16.76 (10)	-4.88 (7)	2.60 (6)	-7.57 (1)	-9.76 (3)
CanCM4 r2 (1)	-17.63 (1)	4.86 (1)	-2.20 (4)	7.11 (7)	-4.00 (2)
CanCM4 r9 (10)	-7.45 (3)	11.7 (6)	5.26 (10)	2.56 (5)	-40.85 (10)
CanCM4 r8 (8)	-23.16 (2)	9.41 (3)	5.01 (9)	10.55 (8)	-35.96 (9)
CanCM4 r6 (9)	14.32 (8)	-8.41 (10)	-3.62 (2)	12.09 (9)	-14.75 (5)
CanCM4 r1 (5)	16.26 (9)	-7.09 (9)	0.91 (5)	-0.51 (2)	-10.9 (4)
FGOALS-g2 r2 (3)	0.55 (3)	-6.75 (2)	-2.63 (1)	11.24 (3)	-5.95 (2)
FGOALS-g2 r3 (2)	-1.63 (2)	-9.48 (3)	-6.53 (2)	9.17 (2)	21.84 (1)
FGOALS-g2 r1 (1)	-29.65 (1)	9.53 (1)	6.06 (3)	3.34 (1)	-18.17 (3)
MIROC4h r2 (1)	-26.62 (1)	-0.18 (1)	-1.60 (1)	18.65 (2)	0.51 (1)
MIROC4h r1 (2)	2.81 (3)	-1.84 (2)	4.01 (3)	3.76 (1)	-22.70 (3)
MIROC4h r3 (3)	-4.31 (2)	-17.45 (3)	-1.40 (2)	23.18 (3)	-4.20 (2)
MIROC5 r2 (4)	-17.69 (1)	-4.81 (5)	-0.95 (2)	8.96 (6)	8.70 (4)
MIROC5 r1 (2)	-5.40 (4)	-1.89 (4)	2.39 (4)	-8.98 (1)	13.74 (2)
MIROC5 r6 (5)	-8.28 (3)	0.36 (2)	4.09 (6)	6.37 (4)	-22.17 (5)
MIROC5 r3 (1)	-20.44 (2)	7.44 (1)	1.28 (3)	-6.87 (2)	13.20 (3)
MIROC5 r5 (3)	13.34 (6)	-9.60 (6)	-4.83 (1)	-2.38 (3)	19.96 (1)
MIROC5 r4 (6)	-1.62 (5)	-0.56 (3)	3.24 (5)	8.94 (5)	-24.20 (6)
MPI-ESM-LR r2 (3)	-34.18 (3)	11.24 (2)	4.64 (3)	12.94 (3)	-24.74 (3)
MPI-ESM-LR r3 (1)	-21.01 (2)	7.54 (1)	-0.82 (1)	0.04 (2)	6.61 (1)
MPI-ESM-LR r1 (2)	-12.27 (1)	15.34 (3)	2.52 (2)	-9.80 (1)	-12.80 (2)
MRI-CGCM3 r3 (1)	-23.64 (1)	5.84 (1)	-0.17 (1)	-6.65 (1)	16.67 (1)
MRI-CGCM3 r1 (3)	-8.94 (2)	0.98 (3)	2.68 (2)	0.60 (3)	-10.97 (3)
MRI-CGCM3 r2 (2)	-40.43 (3)	6.30 (2)	2.73 (3)	0.06 (2)	16.30 (2)

Table A5: As Table A1, but for CMIP3 trends in 1957 – 1978.

SLP (Pa yr^{-1})	80°S-50°S	50°S-20°S	20°S-20°N	20°N-50°N	50°N-80°N
HadSLP2	4.63	0.51	0.05	0.17	-3.24
GISS-ER r3 (1)	5.23 (1)	-3.31 (4)	0.48 (1)	-0.33 (1)	-3.17 (1)
GISS-ER r2 (2)	2.13 (3)	-0.83 (2)	0.86 (3)	-0.53 (2)	-4.85 (3)
GISS-ER r1 (3)	-5.64 (4)	1.46 (1)	1.19 (4)	2.51 (4)	-5.42 (4)
GISS-ER r5 (4)	-7.66 (5)	3.64 (3)	0.68 (2)	3.24 (5)	-4.44 (2)
GISS-ER r4 (5)	6.04 (2)	-4.80 (5)	-1.62 (5)	2.02 (3)	1.49 (5)
MIROC3.2(medres) r3 (1)	-1.61 (3)	-0.21 (1)	0.34 (2)	1.53 (1)	-2.58 (1)
MIROC3.2(medres) r2 (2)	5.00 (1)	-1.64 (2)	0.07 (1)	-3.72 (2)	3.83 (2)
MIROC3.2(medres) r1 (3)	6.71 (2)	-4.08 (3)	0.77 (3)	5.28 (3)	-11.73 (3)
GISS-EH r2 (1)	4.32 (1)	-0.00 (1)	0.70 (1)	-1.13 (1)	-5.13 (2)
GISS-EH r3 (2)	-5.16 (2)	-0.70 (2)	-0.95 (2)	2.26 (2)	6.58 (3)
GISS-EH r1 (3)	-12.10 (3)	5.45 (3)	1.50 (3)	2.45 (3)	-2.46 (1)
MRI-CGCM2.3.2 r4 (1)	-4.84 (4)	1.20 (2)	-0.26 (1)	1.22 (1)	0.41 (2)
MRI-CGCM2.3.2 r3 (2)	-3.76 (2)	0.23 (1)	0.47 (4)	-1.64 (2)	3.95 (4)
MRI-CGCM2.3.2 r5 (3)	-3.96 (3)	-1.55 (5)	-0.35 (2)	5.57 (4)	-2.36 (1)
MRI-CGCM2.3.2 r1 (4)	-14.99 (5)	1.74 (3)	-0.37 (3)	4.65 (3)	2.98 (3)
MRI-CGCM2.3.2 r2 (5)	0.38 (1)	-0.89 (4)	0.47 (5)	-5.75 (5)	9.26 (5)
FGOALS-g1.0 r2 (1)	0.76 (2)	0.72 (1)	-0.42 (1)	-1.12 (2)	-0.92 (1)
FGOALS-g1.0 r1 (2)	1.28 (1)	-0.07 (2)	0.94 (2)	-0.50 (1)	-5.63 (2)
FGOALS-g1.0 r3 (3)	-3.16 (3)	1.84 (3)	-0.89 (3)	-1.22 (3)	2.89 (3)
ECHAM5/MPI-OM r1 (1)	1.26 (1)	-0.82 (3)	-0.11 (2)	3.14 (2)	-6.54 (3)
ECHAM5/MPI-OM r4 (2)	-2.65 (3)	-2.26 (4)	0.08 (1)	3.74 (3)	-1.43 (1)
ECHAM5/MPI-OM r3 (3)	-4.96 (4)	0.70 (1)	-0.37 (3)	-2.14 (1)	9.01 (4)
ECHAM5/MPI-OM r2 (4)	-0.37 (2)	0.32 (2)	-0.73 (4)	4.81 (4)	-5.90 (2)
CGCM3.1(T47) r3 (1)	-5.29 (3)	-0.26 (1)	1.42 (4)	1.55 (2)	-3.27 (1)
CGCM3.1(T47) r1 (2)	11.97 (1)	-2.98 (4)	-0.17 (2)	-0.25 (1)	-9.87 (4)
CGCM3.1(T47) r4 (3)	-15.56 (4)	3.14 (3)	-0.07 (1)	3.26 (5)	2.61 (3)
CGCM3.1(T47) r5 (4)	-4.28 (2)	2.53 (2)	0.83 (3)	2.95 (4)	-10.05 (5)
CGCM3.1(T47) r2 (5)	-15.62 (5)	5.96 (5)	1.78 (5)	2.48 (3)	-6.46 (2)
CCSM3 r5 (1)	-0.06 (1)	-0.27 (2)	-0.08 (1)	-0.57 (1)	1.47 (1)
CCSM3 r7 (2)	-0.10 (3)	1.20 (1)	0.30 (2)	-3.90 (5)	3.77 (4)
CCSM3 r6 (3)	-4.27 (6)	2.71 (4)	1.03 (4)	-3.80 (4)	3.10 (2)
CCSM3 r1 (4)	-0.068 (2)	-3.46 (6)	-2.01 (6)	1.58 (3)	8.66 (6)
CCSM3 r9 (5)	-6.24 (7)	4.05(5)	2.73 (7)	0.91 (2)	-10.13 (3)
CCSM3 r3 (6)	-3.16 (5)	1.46 (3)	1.45 (5)	5.24 (6)	-14.65 (5)
CCSM3 r2 (7)	12.36 (4)	-4.90 (7)	0.66 (3)	5.48 (7)	-15.25 (7)
GFDL-CM2.1 r3 (1)	-4.52 (1)	-1.18 (1)	-0.59 (2)	0.52 (2)	7.30 (3)
GFDL-CM2.1 r1 (2)	-8.35 (2)	3.73 (2)	-0.19 (1)	-0.66 (3)	3.28 (2)
GFDL-CM2.1 r2 (3)	-10.54 (3)	4.97 (3)	1.33 (3)	0.35 (1)	-2.76 (1)

Table A6: As Table A2, but for CMIP3 trends in 1979 – 1989.

SLP (Pa yr^{-1})	80°S-50°S	50°S-20°S	20°S-20°N	20°N-50°N	50°N-80°N
HadSLP2	-9.65	-0.86	3.45	6.93	-21.48
GISS-ER r3 (1)	-14.39 (4)	10.57 (5)	3.42 (1)	5.95 (1)	-19.24 (1)
GISS-ER r2 (5)	-14.56 (5)	2.93 (2)	-0.23 (5)	10.23 (2)	-8.96 (4)
GISS-ER r1 (3)	-8.61 (1)	7.12 (4)	5.81 (4)	-3.09 (4)	-16.62 (2)
GISS-ER r5 (4)	-13.56 (2)	1.00 (1)	2.18 (3)	-3.91 (5)	13.54 (5)
GISS-ER r4 (2)	-14.22 (3)	6.44 (3)	4.48 (2)	3.39 (3)	-15.95 (3)
MIROC3.2(medres) r3 (3)	-3.97 (1)	-4.44 (3)	-1.74 (3)	-0.41 (2)	14.50 (3)
MIROC3.2(medres) r2 (2)	15.91 (3)	-0.72 (1)	1.16 (2)	-10.23 (3)	-2.76 (2)
MIROC3.2(medres) r1 (1)	-24.22 (2)	1.60 (2)	2.45 (1)	7.37 (1)	-4.01 (1)
GISS-EH r2 (1)	-5.45 (1)	2.30 (2)	3.22 (1)	14.29 (3)	-35.30 (2)
GISS-EH r3 (2)	-0.90 (3)	-7.71 (3)	2.38 (2)	3.55 (1)	-11.69 (1)
GISS-EH r1 (3)	-14.99 (2)	1.35 (1)	0.58 (3)	3.46 (2)	2.75 (3)
MRI-CGCM2.3.2 r4 (5)	-19.57 (3)	8.10 (5)	7.81 (5)	-14.45 (5)	4.33 (5)
MRI-CGCM2.3.2 r3 (3)	-25.92 (4)	3.36 (3)	7.77 (4)	1.48 (2)	-7.83 (2)
MRI-CGCM2.3.2 r5 (1)	-4.07 (1)	3.18 (2)	3.37 (1)	-1.49 (3)	-12.06 (1)
MRI-CGCM2.3.2 r1 (2)	-18.98 (2)	2.56 (1)	0.65 (3)	7.16 (1)	0.12 (4)
MRI-CGCM2.3.2 r2 (4)	13.89 (5)	-5.72 (4)	1.17 (2)	-2.41 (4)	-4.40 (3)
FGOALS-g1.0 r2 (2)	-26.86 (2)	8.07 (2)	4.92 (2)	1.01 (2)	-8.78 (2)
FGOALS-g1.0 r1 (1)	-6.33 (1)	1.82 (1)	2.35 (1)	-13.78 (3)	15.85 (3)
FGOALS-g1.0 r3 (3)	-29.22 (3)	9.69 (3)	5.26 (3)	6.76 (1)	-22.18 (1)
ECHAM5/MPI-OM r1 (2)	-11.30 (1)	11.33 (4)	0.55 (2)	-5.24 (3)	-1.44 (2)
ECHAM5/MPI-OM r4 (4)	-16.82 (2)	5.98 (3)	2.44 (1)	-6.49 (4)	9.60 (4)
ECHAM5/MPI-OM r3 (3)	-23.05 (3)	5.53 (2)	-1.05 (4)	8.14 (1)	2.29 (3)
ECHAM5/MPI-OM r2 (1)	14.48 (4)	0.04 (1)	0.24 (3)	4.53 (2)	-28.44 (1)
CGCM3.1(T47) r3 (2)	0.047 (3)	3.83 (1)	3.23 (1)	-17.28 (5)	7.02 (4)
CGCM3.1(T47) r1 (3)	-11.49 (1)	4.11 (2)	2.30 (3)	-12.84 (4)	9.74 (5)
CGCM3.1(T47) r4 (5)	8.08 (5)	-10.24 (5)	1.37 (4)	-0.49 (2)	0.11 (2)
CGCM3.1(T47) r5 (1)	-23.96 (4)	5.90 (4)	3.80 (2)	2.97 (1)	-8.79 (1)
CGCM3.1(T47) r2 (4)	-0.12 (2)	-6.03 (3)	0.63 (5)	-4.21 (3)	6.74 (3)
CCSM3 r5 (3)	-6.50 (2)	1.80 (2)	0.33 (5)	-0.08 (5)	1.55 (5)
CCSM3 r7 (2)	-11.95 (1)	-7.10 (4)	0.36 (4)	2.81 (3)	18.60 (7)
CCSM3 r6 (7)	-21.26 (4)	6.53 (5)	7.69 (7)	13.28 (4)	-38.26 (2)
CCSM3 r1 (4)	-4.27 (3)	0.78 (1)	0.57 (3)	-5.28 (7)	1.94 (6)
CCSM3 r9 (1)	-37.54 (7)	3.00 (3)	3.78 (1)	10.33 (2)	0.26 (4)
CCSM3 r3 (5)	-25.85 (5)	7.96 (6)	0.20 (6)	5.65 (1)	-0.14 (3)
CCSM3 r2 (6)	-34.49 (6)	16.11 (7)	5.03 (2)	-2.74 (6)	-9.15 (1)
GFDL-CM2.1 r3 (1)	-4.46 (2)	-0.59 (1)	2.51 (1)	8.28 (1)	-17.30 (1)
GFDL-CM2.1 r1 (2)	-29.66 (3)	6.22 (2)	0.28 (2)	4.92 (2)	4.65 (2)
GFDL-CM2.1 r2 (3)	-14.32 (1)	7.99 (3)	7.73 (3)	20.42 (3)	-62.17 (3)

Table A7: As Table A1, but for CMIP5 decadal simulations (1961-1980).

SLP (Pa yr^{-1})	80°S-50°S	50°S-20°S	20°S-20°N	20°N-50°N	50°N-80°N
HadSLP2	-5.05	0.88	0.01	0.48	-4.34
CanCM4 r7 (1)	0.18 (4)	1.34 (5)	1.01 (4)	-2.2 (5)	-2.81 (1)
CanCM4 r3 (2)	-1.42 (5)	-0.56 (1)	2.53 (7)	-0.76 (3)	-6.14 (7)
CanCM4 r8 (3)	-4.85 (7)	0.73 (3)	0.47 (2)	3.56 (6)	-5.81 (6)
CanCM4 r6 (4)	-6.69 (10)	2.52 (8)	0.83 (3)	0.99 (2)	-3.66 (3)
CanCM4 r10 (5)	3.12 (1)	-1.59 (4)	1.33 (5)	5.18 (7)	-14.69 (10)
CanCM4 r4 (6)	5.51 (2)	0.48 (2)	3 (9)	-9.02 (10)	-0.88 (4)
CanCM4 r9 (7)	-3.42 (6)	2.11 (6)	2.57 (8)	0.69 (1)	-10.73 (8)
CanCM4 r1 (8)	0.62 (3)	2.3 (7)	3.05 (10)	-8.5 (9)	-1.09 (2)
CanCM4 r5 (9)	13.25 (8)	-5.88 (10)	-0.42 (6)	1.49 (4)	-4.64 (5)
CanCM4 r2 (10)	-5.96 (9)	2.59 (9)	0.56 (1)	-5.26 (8)	6.62 (9)
FGOALS-g2 r1 (1)	4.21 (1)	0.51 (1)	0.94 (2)	-2.09 (1)	-2.37 (1)
FGOALS-g2 r2 (2)	-11.76 (3)	3.3 (3)	0.37 (1)	7.76 (3)	-9.65 (2)
FGOALS-g2 r3 (3)	-3.97 (2)	3.15 (2)	1.47 (3)	6.84 (2)	-19.77 (3)
MIROC5 r6 (1)	1.73 (1)	-0.44 (2)	-0.06 (2)	0.27 (1)	-0.87 (1)
MIROC5 r5 (2)	-0.27 (2)	-0.37 (1)	1.4 (3)	-2.25 (4)	0.73 (2)
MIROC5 r2 (3)	-23.35 (6)	-0.02 (3)	0.49 (1)	7.05 (6)	8.1 (3)
MIROC5 r3 (4)	-2.2 (3)	-1.99 (4)	-0.55 (4)	-1.88 (3)	13.38 (6)
MIROC5 r1 (5)	11.81 (4)	-3.3 (5)	3.08 (6)	1.33 (2)	-16.26 (5)
MIROC5 r4 (6)	-4.98 (5)	-3.49 (6)	-1.32 (5)	2.86 (5)	9.17 (4)
CCSM4 r5 (1)	2.06 (1)	-0.37 (1)	-0.86 (3)	3.85 (7)	-5.96 (3)
CCSM4 r6 (2)	0.35 (3)	0.37 (4)	-0.2 (1)	3.62 (5)	-8.05 (7)
CCSM4 r3 (3)	0.95 (2)	2.26 (6)	-1.68 (9)	-1.15 (2)	1.3 (4)
CCSM4 r8 (4)	-4.06 (6)	3.74 (9)	-1.13 (7)	1.35 (1)	-4.34 (2)
CCSM4 r10 (5)	-5.35 (8)	-0.76 (2)	-0.47 (2)	4.57 (9)	2.45 (6)
CCSM4 r9 (6)	-9.18 (10)	2.18 (5)	-1.05 (6)	3.39 (4)	2.43 (5)
CCSM4 r7 (7)	-6.27 (9)	2.42 (7)	-1.28 (8)	3.74 (6)	-0.68 (1)
CCSM4 r4 (8)	-5.34 (7)	0.34 (3)	-2.16 (10)	3.27 (3)	7.48 (10)
CCSM4 r1 (9)	-3.76 (5)	2.86 (8)	-0.91 (4)	-3.83 (8)	6.55 (9)
CCSM4 r2 (10)	-3.72 (4)	5.18 (10)	-1.03 (5)	5.21 (10)	-10.4 (8)
MRI-CGCM3 r3 (1)	-3.31 (2)	1.96 (2)	0.49 (1)	0.18 (2)	-2.35 (1)
MRI-CGCM3 r2 (2)	-1.78 (1)	-0.25 (1)	-1.37 (2)	2.99 (3)	-0.69 (2)
MRI-CGCM3 r1 (3)	-6.11 (3)	3.13 (3)	2.66 (3)	0.23 (1)	-10.02 (3)
MIROC4h r3 (1)	2.03 (1)	-1.38 (1)	0.41 (1)	3 (1)	-7.8 (3)
MIROC4h r2 (2)	-3.78 (2)	-2.16 (2)	-0.04 (3)	4.4 (3)	-1.72 (1)
MIROC4h r1 (3)	-4.91 (3)	4.46 (3)	0.93 (2)	-3.58 (2)	-0.43 (2)
MPI-ESM-LR r3 (1)	11.59 (2)	-1.51 (1)	-0.64 (1)	-0.8 (1)	-6.73 (2)
MPI-ESM-LR r2 (2)	0.15 (1)	1.09 (2)	-1.12 (2)	3.47 (2)	-5.94 (1)
MPI-ESM-LR r1 (3)	14.35 (3)	-4.57 (3)	-2.55 (3)	4.84 (3)	-8.26 (3)

Table A8: As Table A2, but for CMIP5 decadal simulations (1981-1990).

SLP (Pa yr^{-1})	80°S-50°S	50°S-20°S	20°S-20°N	20°N-50°N	50°N-80°N
HadSLP2	-14.637	-4.425	1.6549	13.429	-19.969
CanCM4 r7 (2)	-14.06 (1)	8.99 (8)	-0.37 (3)	8.59 (4)	-19.42 (1)
CanCM4 r3 (1)	-17.27 (2)	2.02 (3)	0.9 (2)	15.24 (2)	-22.34 (2)
CanCM4 r8 (4)	-21.98 (4)	8.02 (7)	3.81 (4)	11.79 (1)	-33.32 (6)
CanCM4 r6 (8)	-30.62 (8)	12.22 (10)	5.07 (7)	-2.21 (7)	-8.83 (5)
CanCM4 r10 (7)	-22.08 (6)	10.64 (9)	3.93 (5)	-6.42 (8)	-4.71 (8)
CanCM4 r4 (5)	-26.59 (7)	4.46 (5)	-0.81 (6)	17.74 (3)	-10.29 (3)
CanCM4 r9 (6)	2.00 (9)	-8.54 (1)	-2.14 (8)	27.7 (6)	-34.31 (7)
CanCM4 r1 (10)	5.50 (10)	0.23 (2)	-2.37 (9)	-6.63 (9)	3.9 (10)
CanCM4 r5 (9)	-11.87 (3)	7.39 (6)	7.29 (10)	-17.74 (10)	1.09 (9)
CanCM4 r2 (3)	-7.21 (5)	2.41 (4)	1.92 (1)	1.17 (5)	-8.93 (4)
FGOALS-g2 r1 (1)	-18.3 (1)	7.34 (2)	6.81 (3)	1.66 (2)	-23.37 (1)
FGOALS-g2 r2 (3)	-72.56 (3)	10.34 (3)	1.22 (1)	12.75 (1)	23.44 (3)
FGOALS-g2 r3 (2)	5.5 (2)	2.39 (1)	-0.77 (2)	-7.72 (3)	-1.22 (2)
MIROC5 r6 (3)	7.16 (5)	4.02 (5)	2.82 (2)	-3.86 (5)	-16.8 (1)
MIROC5 r5 (1)	-14.44 (1)	-5.01 (1)	-2.95 (4)	9.25 (1)	10.66 (4)
MIROC5 r2 (5)	10.55 (6)	-12.08 (3)	-2.11 (3)	-3.2 (4)	11.51 (5)
MIROC5 r3 (6)	-0.29 (4)	5.53 (6)	-3.42 (5)	-5.49 (6)	5.21 (3)
MIROC5 r1 (4)	-0.74 (3)	-9.04 (2)	-6.15 (6)	7.97 (2)	14.5 (6)
MIROC5 r4 (2)	-18.44 (2)	3.76 (4)	1.5 (1)	6.85 (3)	-13.6 (2)
CCSM4 r5 (2)	-7.66 (4)	-4.67 (2)	-3.18 (10)	11.55 (3)	3.41 (4)
CCSM4 r6 (4)	-23.37 (7)	3.52 (7)	0.77 (3)	7.08 (4)	3.61 (5)
CCSM4 r3 (3)	-6.17 (6)	4.16 (8)	2.51 (2)	0.03 (8)	-13.23 (2)
CCSM4 r8 (8)	-7.31 (5)	13.5 (10)	4.06 (7)	11.93 (2)	-51.35 (9)
CCSM4 r10 (7)	-4.88 (8)	-0.7 (3)	0.66 (4)	-2.45 (10)	5.11 (6)
CCSM4 r9 (6)	-16.16 (1)	6.56 (9)	-1.67 (8)	21.92 (6)	-30.33 (3)
CCSM4 r7 (9)	0.88 (9)	-9.95 (5)	-2.34 (9)	5.15 (5)	10.95 (7)
CCSM4 r4 (1)	-12.95 (2)	3.06 (6)	2.02 (1)	14.8 (1)	-25.49 (1)
CCSM4 r1 (10)	1.05 (10)	-4.25 (1)	-0.31 (6)	-2.23 (9)	16.54 (10)
CCSM4 r2 (5)	-19.1 (3)	-0.57 (4)	-0.17 (5)	3.6 (7)	11.24 (8)
MRI-CGCM3 r3 (3)	-1.69 (2)	-1.59 (1)	-3.66 (2)	-10.4 (3)	17.01 (2)
MRI-CGCM3 r2 (1)	-49.35 (3)	12.52 (2)	1.83 (1)	-3.12 (1)	22.6 (3)
MRI-CGCM3 r1 (2)	-3.61 (1)	14.84 (3)	9.94 (3)	-6.55 (2)	-38.41 (1)
MIROC4h r3 (2)	-22.39 (2)	3.31 (2)	2.5 (1)	9.17 (2)	-8.97 (2)
MIROC4h r2 (3)	-7.85 (1)	7.85 (3)	0.2 (2)	-6.62 (3)	5.07 (3)
MIROC4h r1 (1)	-25.33 (3)	2.97 (1)	3.33 (3)	10.02 (1)	-14.5 (1)
MPI-ESM-LR r3 (3)	-30.82 (3)	11.41 (3)	-2.59 (3)	14.3 (1)	-6.66 (1)
MPI-ESM-LR r2 (2)	-25.89 (2)	6.75 (2)	5.78 (2)	17.87 (2)	-43.66 (2)
MPI-ESM-LR r1 (1)	-24.73 (1)	5.76 (1)	2.67 (1)	-2.66 (3)	4.37 (3)

Table A9: Average and standard deviation for all CMIP3 1957 – 1978 ensemble members and all except lowest performing ensemble members.

SLP (Pa yr^{-1})	80°S-50°S	50°S-20°S	20°S-20°N	20°N-50°N	50°N-80°N
HadSLP2	4.63	0.51	0.05	0.17	-3.24
Avg. (St. Dev.) 38 Ens.	-2.47 (7.16)	0.40 (2.97)	0.25 (1.01)	0.96 (3.08)	-1.42 (6.73)
Avg. (St. Dev.) 24: Ens.	-4.11 (6.17)	1.24 (2.45)	0.55 (0.83)	1.03 (2.74)	-2.28 (4.51)

Table A10: As Table A9, but for the CMIP3 1979 – 1989 hindcast period.

SLP (Pa yr^{-1})	80°S-50°S	50°S-20°S	20°S-20°N	20°N-50°N	50°N-80°N
HadSLP2	-9.65	-0.86	3.4463	6.93	-21.48
Avg. (Stdev.) 38 Ens.	-11.90 (16.00)	2.87 (6.08)	2.56 (3.12)	1.45 (9.26)	-6.74 (19.84)
Avg. (Stdev.) 24 Ens.	-10.15 (14.88)	2.57 (6.37)	2.58 (3.11)	-0.94 (9.53)	-3.78 (16.53)

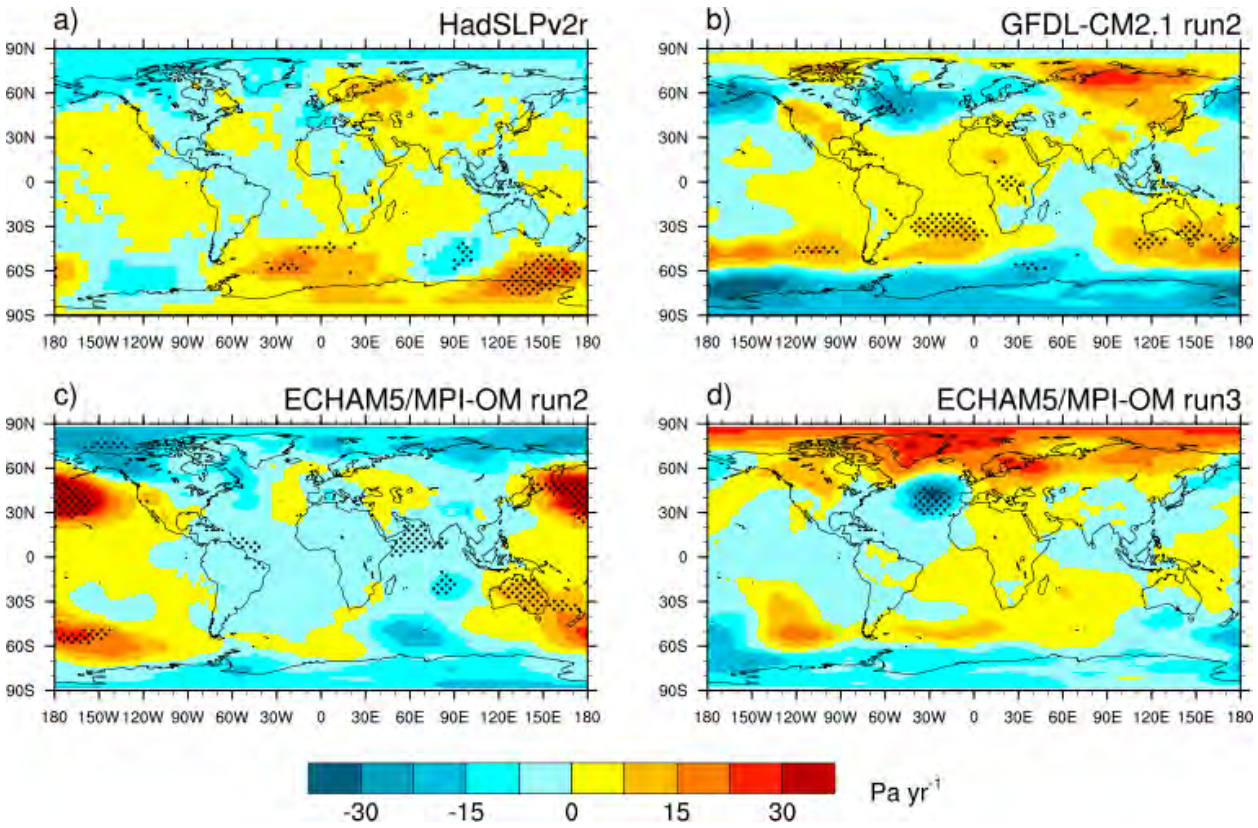

Fig. A1: SLP trends for 1957 – 1978 from a HadSLP2 observations and b-d individual GFDL-CM2.1 and ECHAM5/MPI-OM ensemble members that are found to have zonal trends different from HadSLP2.

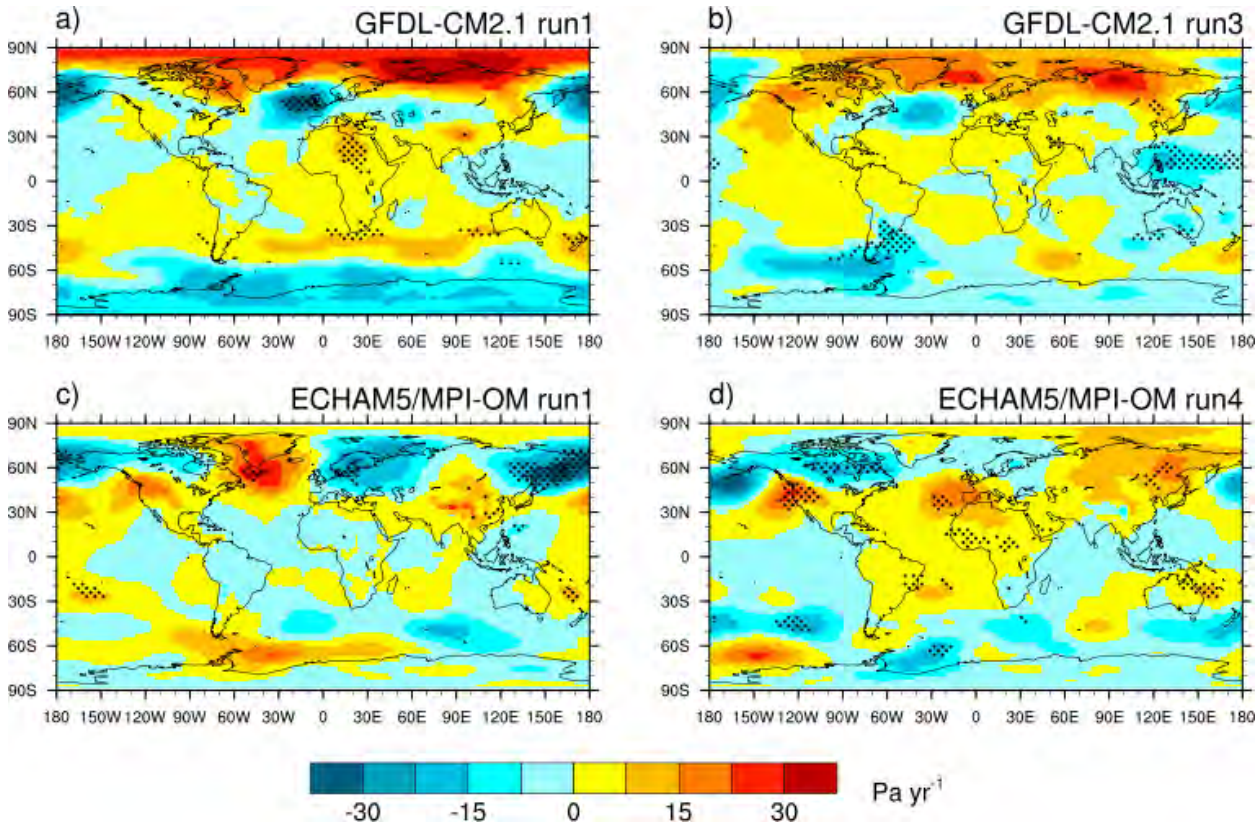

Fig. A2: As Fig. A1, but a-b for individual GFDL-CM2.1 and c-d ECHAM5/MPI-OM ensemble members that are found to have zonal trends similar to HadSLP2.

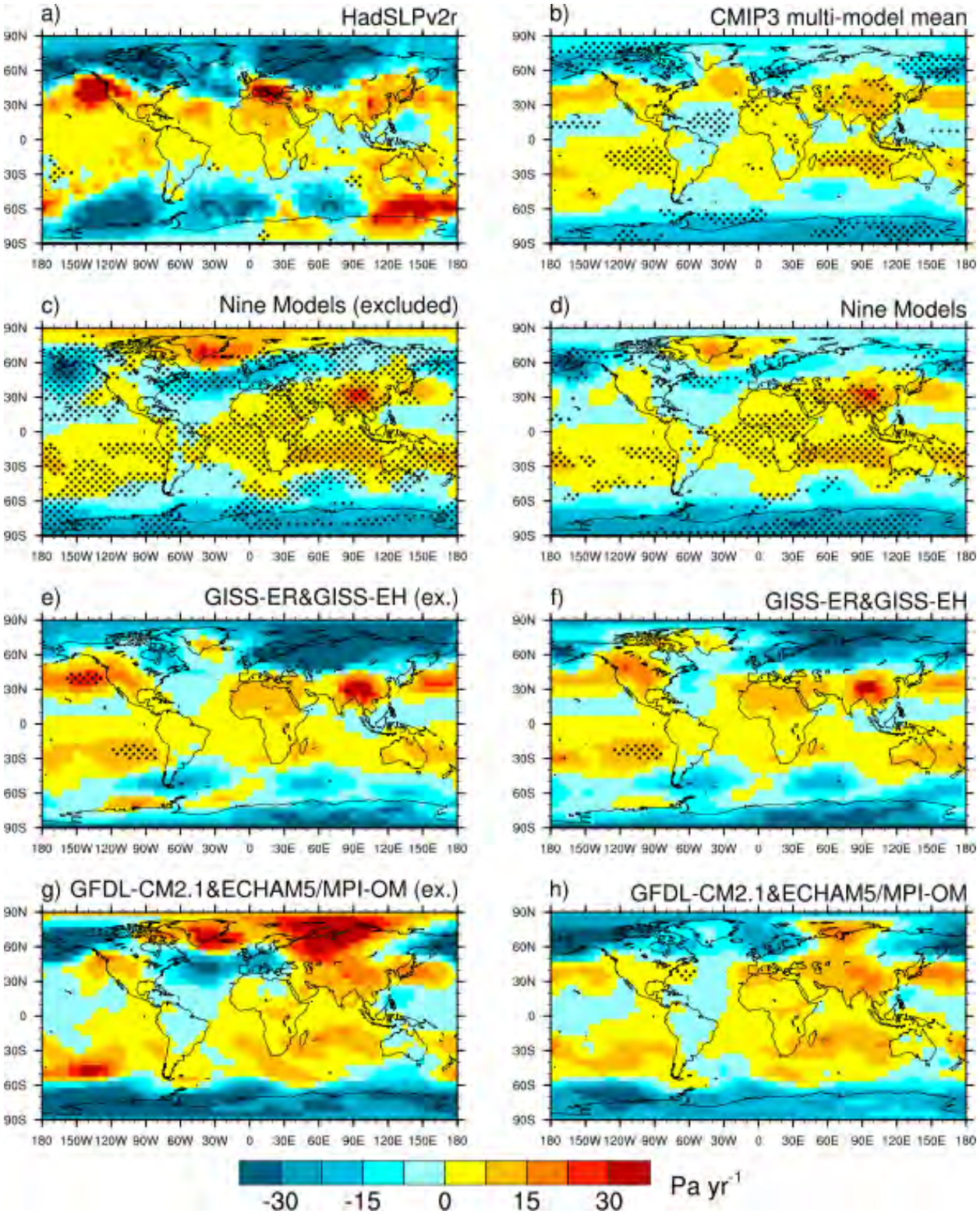

Fig. A3: SLP trends for 1979-1989 from a HadSLP2, b CMIP3 multi-model ensemble, c subset of ensemble members from nine GCMs, d as c but all ensemble members from nine GCMs, e sub-set of ensemble members from GISS-ER and GISS-EH, f as e but all ensemble members from GISS-ER and GISS-EH, g a subset of ensemble members from GFDL-CM2.1 and ECHAM5/ MPI-OM, h as g but all ensemble members from GFDL-CM2.1 and ECHAM5/MPI-OM.

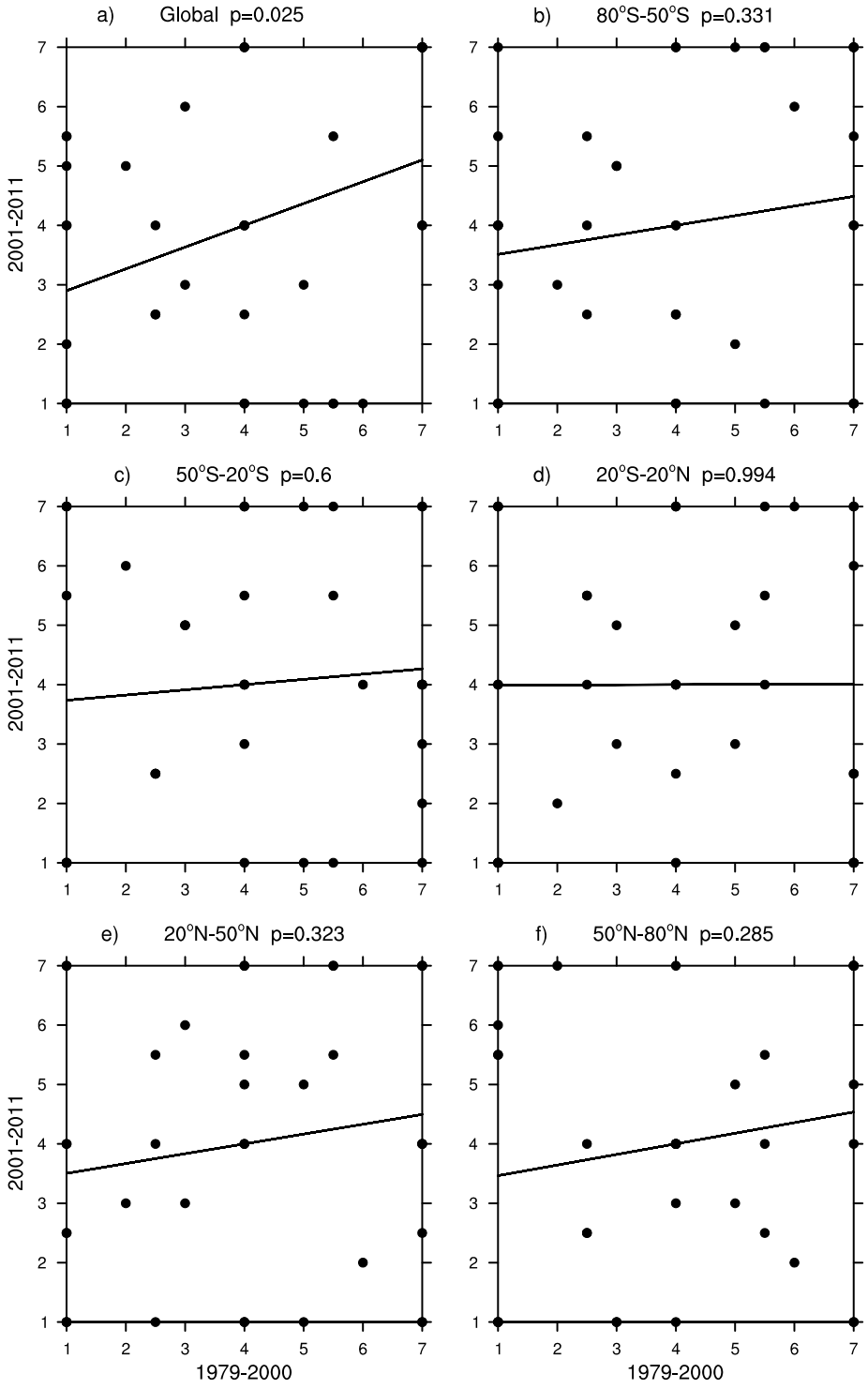

Fig. A4: Regression between the ensemble ranks in training and testing periods for CMIP3 during 1979-2011. *p* values are based on reduced degrees of freedom from multiple GCMs.

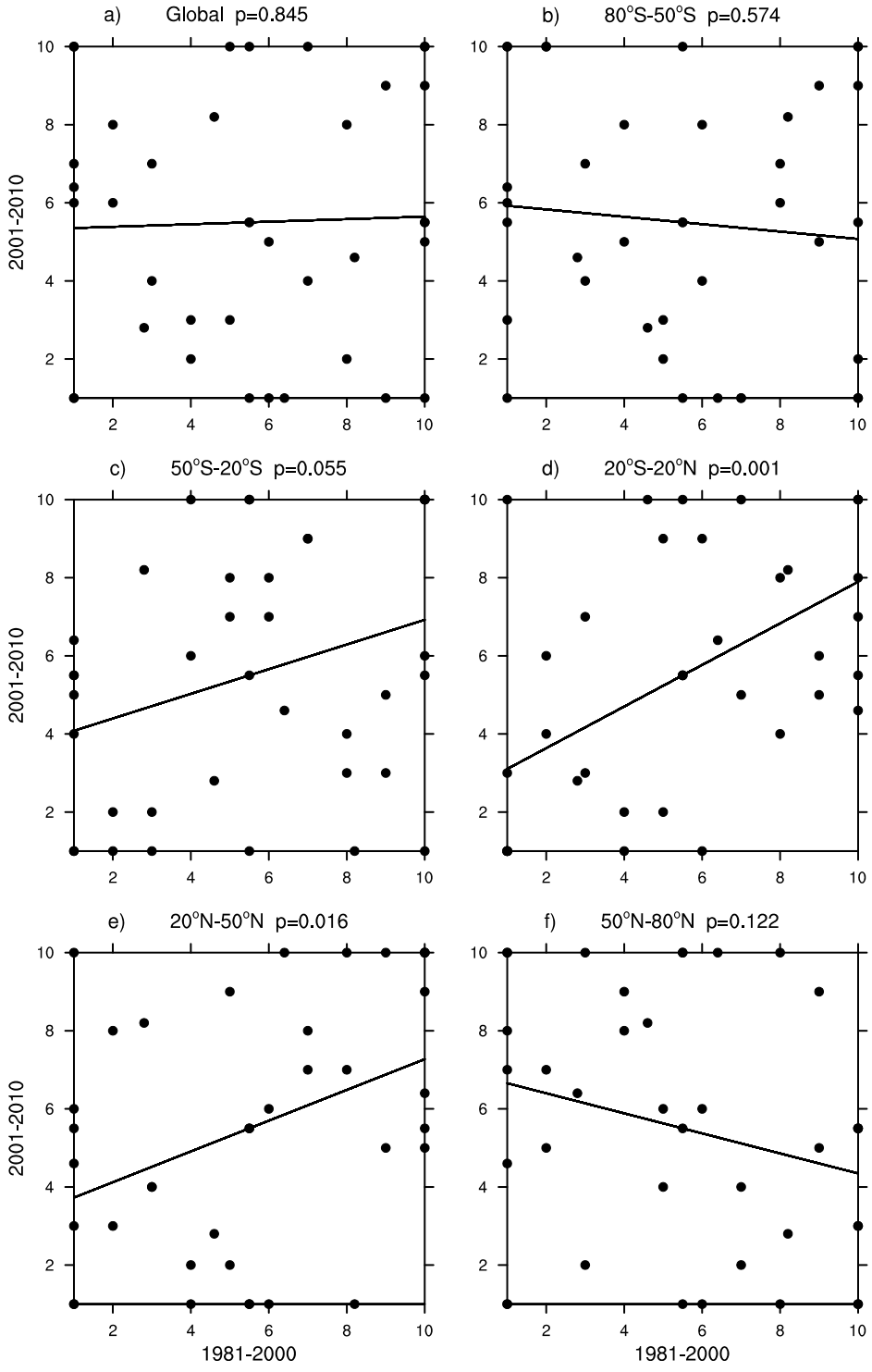

Fig. A5: As for Fig. A4, but for CMIP5 during 1981-2010.

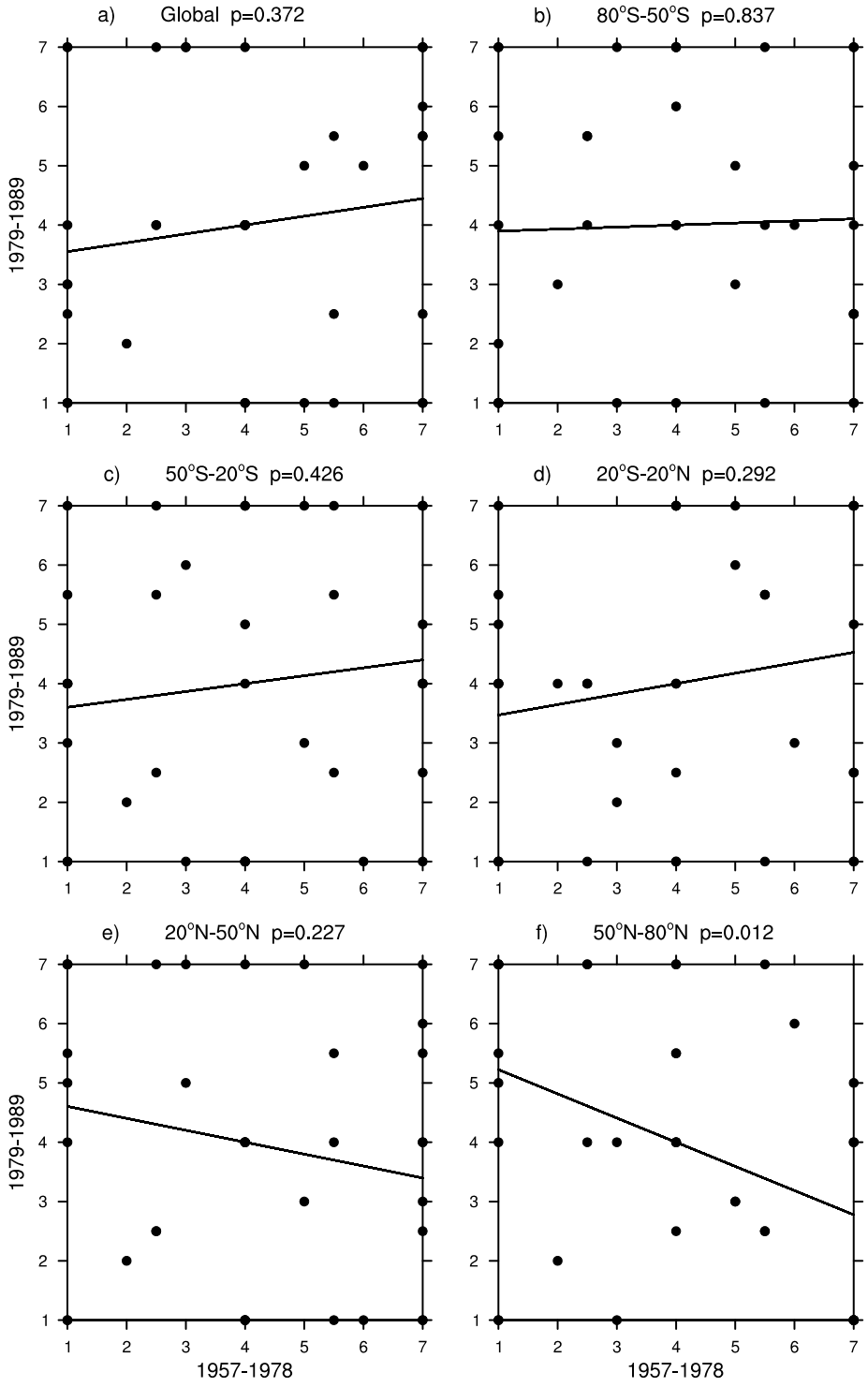

Fig. A6: As for Fig. A4, but for CMIP3 during 1957-1989.

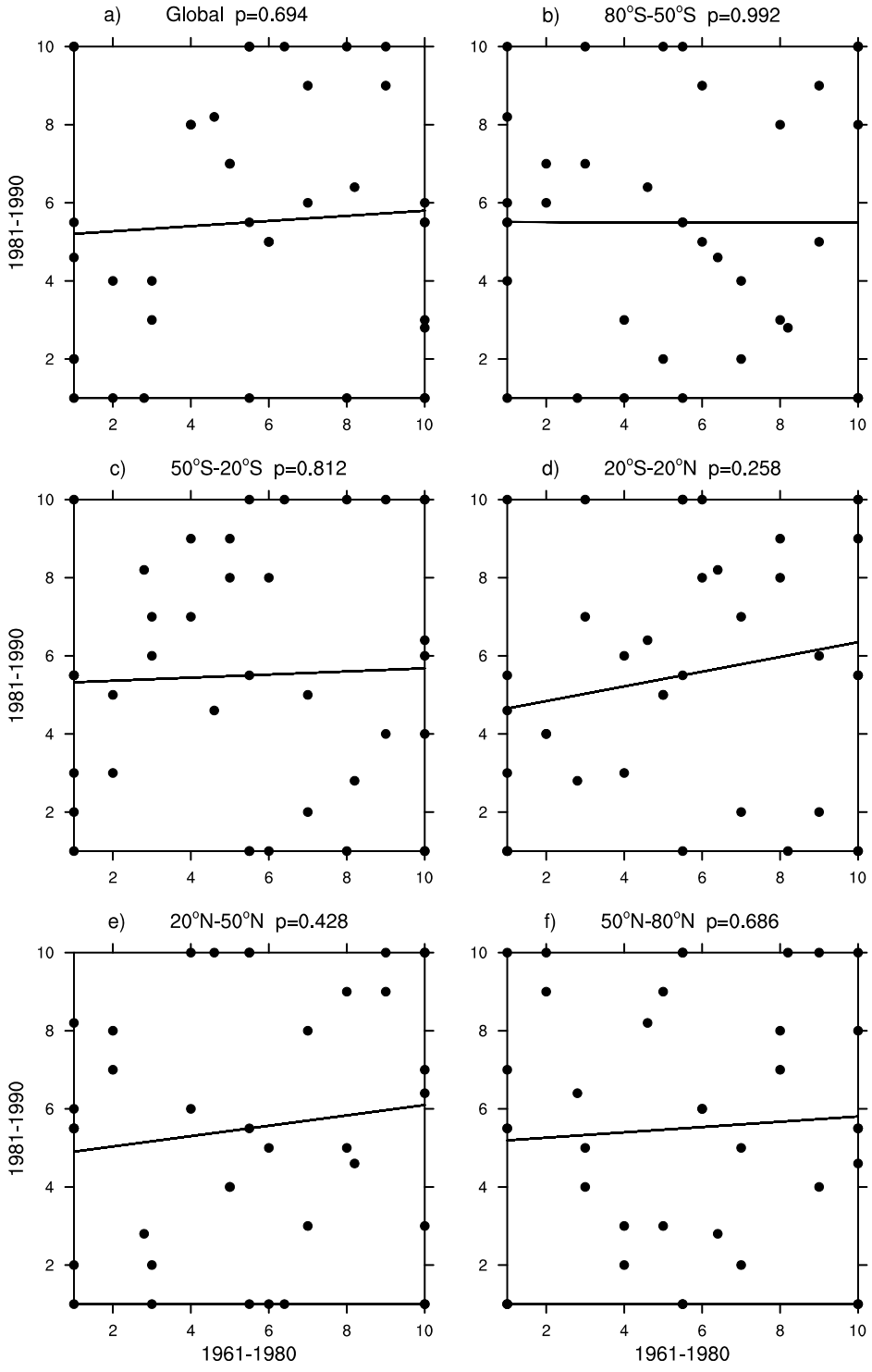

Fig. A7: As for Fig. A4, but for CMIP5 during 1961-1990.

References

Allan R, Ansell T (2006) A new globally complete monthly historical gridded mean sea level pressure dataset (HadSLP2): 1850–2004. J Climate 19**:**5816-5842

Arblaster JM, Meehl GA (2006) Contributions of external forcings to Southern Annular Mode trends. J Climate 19**:**2896-2905

Bellucci A, Haarsma R, Bellouin N, Booth B, Cagnazzo C, van den Hurk B, Keenlyside N, Koenigk T, Massonnet F, Materia S, Weiss M (2015) Advancements in decadal climate predictability: The role of nonoceanic drivers. Reviews of Geophysics 53**:**165-202

Boer GJ, Kharin VV, Merryfield WJ (2013) Decadal predictability and forecast skill. Climate Dyn 41**:**1817-1833

Branstator G (2002) Circumglobal teleconnections, the jet stream waveguide, and the North Atlantic Oscillation. J Climate 15**:**1893-1910

Casado MJ, Pastor MA (2012) Use of variability modes to evaluate AR4 climate models over the Euro-Atlantic region. Climate Dynamics 38**:**225-237

Chylek P, Li J, Dubey M, Wang M, Lesins G (2011) Observed and model simulated 20th century Arctic temperature variability: Canadian Earth System Model CanESM2. Atmos Chem Phys Discuss 11**:**22893–22907

Clement A, Bellomo K, Murphy LN, Cane MA, Mauritsen T, Rädel G, Stevens B (2015) The Atlantic Multidecadal Oscillation without a role for ocean circulation. Science 350**:**320-324

Collins M (2007) Ensembles and probabilities: A new era in the prediction of climate change. Phil Trans Roy Soc 365A**:**1957-1970

Collins WD, Bitz CM, Blackmon ML, Bonan GB, Bretherton CS, Carton JA, Chang P, Doney SC, Hack JJ, Henderson TB, Kiehl JT, Large WG, Mckenna DS, Santer BD, Smith RD (2006) The Community Climate System Model Version 3 (CCSM3). Journal of Climate 19**:**2122-2143

d'Orgeville M, Peltier WR (2007) On the Pacific Decadal Oscillation and the Atlantic Multidecadal Oscillation: Might they be related? Geophys Res Lett 34**:**L23705

DelSole T, Tippett MK, Shukla J (2011) A significant component of unforced multidecadal variability in the recent acceleration of global warming. J Climate 24**:**909-926

Delworth TL, Broccoli AJ, Rosati A, Stouffer RJ, Balaji V, Beesley JA, Cooke WF, Dixon KW, Dunne J, Dunne KA, Durachta JW, Findell KL, Ginoux P, Gnanadesikan A, Gordon CT, Griffies SM, Gudgel R, Harrison MJ, Held IM, Hemler RS, Horowitz LW, Klein SA, Knutson TR, Kushner PJ, Langenhorst AR, Lee H-C, Lin S-J, Lu J, Malyshev SL, Milly PCD, Ramaswamy V, Russell J, Schwarzkopf MD, Shevliakova E, Sirutis JJ, Spelman MJ, Stern WF, Winton M, Wittenberg AT, Wyman B, Zeng F, Zhang R (2006) GFDL's CM2 global coupled climate models. part I: formulation and simulation characteristics. J Climate 19**:**643-674

Delworth TL, Mann ME (2000) Observed and simulated multidecadal variability in the Northern Hemisphere. Climate Dyn 16**:**661-676

Delworth TL, Zhang R, Mann ME (2007) Decadal to centennial variability of the Atlantic from observations and models. In: Ocean Circulation: Mechanisms and Impacts - Past and Future Changes of Meridional Overturning. AGU, Washington, DC, pp 131-148

Denton GH, Broecker WS (2008) Wobbly ocean conveyor circulation during the Holocene? Quaternary Science Reviews 27:1939-1950
Dima M, Lohmann G (2007) A hemispheric mechanism for the Atlantic Multidecadal Oscillation. J Climate 20:2706-2719
Ding Q, Wang B (2005) Circumglobal teleconnection in the Northern Hemisphere summer. J Climate 18:3483-3505
Feldstein SB, Franzke C (2006) Are the North Atlantic Oscillation and the Northern Annular Mode distinguishable? J Atmos Sci 63:2915-2930
Flatau M, Kim Y-J (2013) Interaction between the MJO and Polar Circulations. J Climate 26:3562-3574
Folland CK, Palmer TN, Parker DE (1986) Sahel rainfall and worldwide sea temperatures, 1901-85. Nature 320:602-607
Gent PR, Danabasoglu G, Donner LJ, Holland MM, Hunke EC, Jayne SR, Lawrence DM, Neale RB, Rasch PJ, Vertenstein M, Worley PH, Yang Z-L, Zhang M (2011) The Community Climate System Model version 4. J Climate 24:4973-4991
Gerber EP, Baldwin MP, Akiyoshi H, Austin J, Bekki S, Braesicke P, Butchart N, Chipperfield M, Dameris M, Dhomse S, Frith SM, Garcia RR, Garny H, Gettelman A, Hardiman SC, Karpechko A, Marchand M, Morgenstern O, Nielsen JE, Pawson S, Peter T, Plummer DA, Pyle JA, Rozanov E, Scinocca JF, Shepherd TG, Smale D (2010) Stratosphere-troposphere coupling and annular mode variability in chemistry-climate models. J Geophys Res 115:D00M06
Gerber EP, Polvani LM, Ancukiewicz D (2008) Annular mode time scales in the Intergovernmental Panel on Climate Change Fourth Assessment Report models. Geophysical Research Letters 35:L22707
Gillett NP, Allan RJ, Ansell TJ (2005) Detection of external influence on sea level pressure with a multi-model ensemble. Geophys Res Lett 32:L19714
Gillett NP, Fyfe JC (2013) Annular mode changes in the CMIP5 simulations. Geophys Res Lett 40:1189-1193
Gillett NP, Fyfe JC, Parker DE (2013) Attribution of observed sea level pressure trends to greenhouse gas, aerosol, and ozone changes. Geophys Res Lett 40:2302-2306
Gillett NP, Stott PA (2009) Attribution of anthropogenic influence on seasonal sea level pressure. Geophys Res Lett 36:L23709, doi: 23710.21029/22009gl041269
Gillett NP, Thompson DWJ (2003) Simulation of recent Southern Hemisphere climate change. Science 302:273-275
Gillett NP, Zwiers FW, Weaver AJ, Stott PA (2003) Detection of human influence on sea-level pressure. Nature 422:292-294
Gleckler PJ, Taylor KE, Doutriaux C (2008) Performance metrics for climate models. J Geophys Res 113:D06104
Handorf D, Dethloff K (2012) How well do state-of-the-art atmosphere-ocean general circulation models reproduce atmospheric teleconnection patterns? Tellus A: Dynamic Meteorology and Oceanography 64:19777
Hegerl GC, Hasselmann K, Cubasch U, Mitchell JFB, Roeckner E, Voss R, Waszkewitz J (1997) Multi-fingerprint detection and attribution analysis of greenhouse gas, greenhouse gas-plus-aerosol and solar forced climate change. Climate Dyn 13:613-634
Hurrell JW (1995) Decadal trends in the North Atlantic Oscillation: regional temperatures and precipitation. Science 269:676-679

Hurrell JW (1996) Influence of variations in extratropical wintertime teleconnections on Northern Hemisphere temperature. Geophys Res Lett 23:665-668
Karpechko AY, Gillett NP, Marshall GJ, Screen JA (2009) Climate impacts of the Southern Annular Mode simulated by the CMIP3 models. J Climate 22:3751-3768
Keenlyside NS, Latif M, Jungclaus J, Kornblueh L, Roeckner E (2008) Advancing decadal-scale climate prediction in the North Atlantic sector. Nature 453:84-88
Kim H-M, Webster PJ, Curry JA (2012) Evaluation of short-term climate change prediction in multi-model CMIP5 decadal hindcasts. Geophys Res Lett 39:L10701
Knight JR, Allan RJ, Folland CK, Vellinga M, Mann ME (2005) A signature of persistent natural thermohaline circulation cycles in observed climate. Geophys Res Lett 32:L20708
Knutti R, Sedlacek J (2013) Robustness and uncertainties in the new CMIP5 climate model projections. Nature Clim Change 3:369-373
L'Heureux ML, Thompson DWJ (2006) Observed relationships between the El Niño–Southern Oscillation and the extratropical zonal-mean circulation. J Climate 19:276-287
Laepple T, Jewson S, Coughlin K (2008) Interannual temperature predictions using the CMIP3 multi-model ensemble mean. Geophys Res Lett 35:L10701
Latif M, Arpe K, Roeckner E (2000) Oceanic control of decadal North Atlantic sea level pressure variability in winter. Geophysical Research Letters 27:727-730
Lee M-Y, Hsu H-H (2013) Identification of the Eurasian–North Pacific Multidecadal Oscillation and its relationship to the AMO. J Climate 26:8139-8153
Li L, Lin P, Yu Y, Wang B, Zhou T, Liu L, Liu J, Bao Q, Xu S, Huang W, Xia K, Pu Y, Dong L, Shen S, Liu Y, Hu N, Liu M, Sun W, Shi X, Zheng W, Wu B, Song M, Liu H, Zhang X, Wu G, Xue W, Huang X, Yang G, Song Z, Qiao F (2013) The flexible global ocean-atmosphere-land system model, Grid-point Version 2: FGOALS-g2. Advances in Atmospheric Sciences 30:543-560
Liess S, Agrawal S, Chatterjee S, Kumar V (2017) A teleconnection between the West Siberian Plain and the ENSO region. J Climate 30:301-315
Liess S, Kumar A, Snyder PK, Kawale J, Steinhaeuser K, Semazzi FHM, Ganguly AR, Samatova NF, Kumar V (2014) Different modes of variability over the Tasman Sea: Implications for regional climate. J Climate 27:8466-8486
Mahajan S, Zhang R, Delworth TL (2011a) Impact of the Atlantic Meridional Overturning Circulation (AMOC) on Arctic surface air temperature and sea ice variability. J Climate 24:6573-6581
Mahajan S, Zhang R, Delworth TL, Zhang S, Rosati AJ, Chang Y-S (2011b) Predicting Atlantic Meridional Overturning Circulation (AMOC) variations using subsurface and surface fingerprints. Deep Sea Research Part II: Topical Studies in Oceanography 58:1895-1903
Matei D, Baehr J, Jungclaus JH, Haak H, Müller WA, Marotzke J (2012) Multiyear Prediction of Monthly Mean Atlantic Meridional Overturning Circulation at 26.5°N. Science 335:76-79
McFarlane N, Scinocca J, Lazare M, Harvey R, Verseghy D, Li J (2005) The CCCma third generation atmospheric general circulation model. CCCma internal rep 25:16
Medhaug I, Furevik T (2011) North Atlantic 20th century multidecadal variability in coupled climate models: sea surface temperature and ocean overturning circulation. Ocean Sci 7:389-404

Meehl GA, Goddard L, Murphy J, Stouffer RJ, Boer G, Danabasoglu G, Dixon K, Giorgetta MA, Greene AM, Hawkins E, Hegerl G, Karoly D, Keenlyside N, Kimoto M, Kirtman B, Navarra A, Pulwarty R, Smith D, Stammer D, Stockdale T (2009) Decadal prediction: Can it be skillful? Bull Amer Meteor Soc 90**:**1467-1485

Miller RL, Schmidt GA, Shindell DT (2006) Forced annular variations in the 20th century Intergovernmental Panel on Climate Change Fourth Assessment Report models. J Geophys Res 111**:**D18101

Mosteller F, Fisher RA (1948) Questions and answers: combining independent tests of significance. The American Statistician 2**:**30-31

Nakicenovic N, Alcamo J, Davis G, de Vries B, Fenhann J, Gaffin S, Gregory K, Grubler A, Jung TY, Kram T (2000) Emissions Scenarios: Summary for Policymakers. In: Nakicenovic N, Swart R (Eds.) Special Report on Emissions Scenarios: A Special Report of Working Group III of the Intergovernmental Panel on Climate Change. Intergovernmental Panel on Climate Change, 27

NASA (1976) US Standard Atmosphere. National Aeronautics and Space Administration, United States Air Force, Washington, DC**:**243

Nozawa T, Nagashima T, Tomo'o Ogura TY, Okada N, Shiogama H (2007) Climate change simulations with a coupled ocean-atmosphere GCM called the Model for Interdisciplinary Research on Climate: MIROC, Cent. For Global Environ. Res., Natl. Inst. for Environ. Stud., Tsukuba, Japan, pp 80

Pastor MA, Casado MJ (2012) Use of circulation types classifications to evaluate AR4 climate models over the Euro-Atlantic region. Climate Dynamics 39**:**2059-2077

Philipp A, Della-Marta PM, Jacobeit J, Fereday DR, Jones PD, Moberg A, Wanner H (2007) Long-term variability of daily North Atlantic–European pressure patterns since 1850 classified by simulated annealing clustering. J Climate 20**:**4065-4095

Pincus R, Batstone CP, Hofmann RJP, Taylor KE, Glecker PJ (2008) Evaluating the present-day simulation of clouds, precipitation, and radiation in climate models. J Geophys Res 113**:**D14209

Pohlmann H, Jungclaus JH, Köhl A, Stammer D, Marotzke J (2009) Initializing decadal climate predictions with the GECCO oceanic synthesis: Effects on the North Atlantic. J Climate 22**:**3926-3938

Raddatz T, Reick C, Knorr W, Kattge J, Roeckner E, Schnur R, Schnitzler KG, Wetzel P, Jungclaus J (2007) Will the tropical land biosphere dominate the climate–carbon cycle feedback during the twenty-first century? Climate Dyn 29**:**565-574

Roeckner E, Baeuml G, Bonaventura L, Brokopf R, Esch M, Giorgetta M, Hagemann S, Kirchner I, Kornblueh L, Manzini E, Rhodin A, Schlese U, Schulzweida U, Tompkins A (2003) The atmospheric general circulation model ECHAM5 Part I: Model description. Report 349, Max-Planck-Institut fuer Meteorologie, Hamburg, Germany, pp 140

Ruiz-Barradas A, Nigam S, Kavvada A (2013) The Atlantic Multidecadal Oscillation in twentieth century climate simulations: uneven progress from CMIP3 to CMIP5. Climate Dyn 41**:**3301-3315

Sakamoto T, Komuro Y, Ishii M, Tatebe H, Shiogama H, Hasegawa A, Toyoda T, Mori M, Suzuki T, Imada-Kanamaru Y (2012) MIROC4h—a new high-resolution atmosphere-ocean coupled general circulation model. J Meteorol Soc Jpn 90**:**325-359

Santer BD, Wigley TML, Boyle JS, Gaffen DJ, Hnilo JJ, Nychka D, Parker DE, Taylor KE (2000) Statistical significance of trends and trend differences in layer-average atmospheric temperature time series. J Geophys Res 105:7337-7356

Schmidt GA, Ruedy R, Hansen JE, Aleinov I, Bell N, Bauer M, Bauer S, Cairns B, Canuto V, Cheng Y, Del Genio A, Faluvegi G, Friend AD, Hall TM, Hu Y, Kelley M, Kiang NY, Koch D, Lacis AA, Lerner J, Lo KK, Miller RL, Nazarenko L, Oinas V, Perlwitz J, Perlwitz J, Rind D, Romanou A, Russell GL, Sato M, Shindell DT, Stone PH, Sun S, Tausnev N, Thresher D, Yao M-S (2006) Present-day atmospheric simulations using GISS ModelE: comparison to in situ, satellite, and reanalysis data. J Climate 19:153-192

Schubert S, Wang H, Suarez M (2011) Warm season subseasonal variability and climate extremes in the Northern Hemisphere: the role of stationary Rossby waves. J Climate 24:4773-4792

Simmonds I (2015) Comparing and contrasting the behaviour of Arctic and Antarctic sea ice over the 35 year period 1979–2013. Annals of Glaciology 56:18-28

Stoner AMK, Hayhoe K, Wuebbles DJ (2009) Assessing general circulation model simulations of atmospheric teleconnection patterns. J Climate 22:4348-4372

Sutton RT, Hodson DLR (2005) Atlantic ocean forcing of North American and European summer climate. Science 309:115-118

Teng H, Branstator G, Wang H, Meehl GA, Washington WM (2013) Probability of US heat waves affected by a subseasonal planetary wave pattern. Nature Geosci 6:1056-1061

Thompson DWJ, Wallace JM (2001) Regional climate impacts of the Northern Hemisphere Annular Mode. Science 293:85-89

Trenberth KE (1981) Seasonal variations in global sea level pressure and the total mass of the atmosphere. J Geophys Res 86:5238-5246

Trenberth KE (1991) Storm tracks in the Southern Hemisphere. J Atmos Sci 48:2159-2178

Trenberth KE, Stepaniak DP, Smith L (2005) Interannual variability of patterns of atmospheric mass distribution. J Climate 18:2812-2825

van Oldenborgh GJ, cited 2016: KNMI Climate Explorer. [Available online at https://climexp.knmi.nl.]

van Oldenborgh GJ, Doblas-Reyes F, Wouters B, Hazeleger W (2012) Decadal prediction skill in a multi-model ensemble. Climate Dyn 38:1263-1280

van Oldenborgh GJ, te Raa LA, Dijkstra HA, Philip SY (2009) Frequency- or amplitude-dependent effects of the Atlantic meridional overturning on the tropical Pacific Ocean. Ocean Sci 5:293-301

van Vuuren DP, Edmonds J, Kainuma M, Riahi K, Thomson A, Hibbard K, Hurtt GC, Kram T, Krey V, Lamarque J-F, Masui T, Meinshausen M, Nakicenovic N, Smith SJ, Rose SK (2011) The representative concentration pathways: an overview. Climatic Change 109:5-31

Wallace JM (2000) North Atlantic Oscillation/annular mode: Two paradigms—one phenomenon. Quart J Roy Meteor Soc 126:791-805

Walsh JE, Chapman WL, Shy TL (1996) Recent decrease of sea level pressure in the central Arctic. J Climate 9:480-486

Watanabe M, Suzuki T, O'ishi R, Komuro Y, Watanabe S, Emori S, Takemura T, Chikira M, Ogura T, Sekiguchi M, Takata K, Yamazaki D, Yokohata T, Nozawa T,

Hasumi H, Tatebe H, Kimoto M (2010) Improved climate simulation by MIROC5: mean states, variability, and climate sensitivity. J Climate 23**:**6312-6335

Wilmes SB, Raible CC, Stocker TF (2012) Climate variability of the mid- and high-latitudes of the Southern Hemisphere in ensemble simulations from 1500 to 2000 AD. Clim Past 8**:**373-390

Xin X-G, Zhou T-J, Yu R-C (2008) The Arctic Oscillation in coupled climate models. Chinese Journal of Geophysics 51**:**223-239

Yeager S, Karspeck A, Danabasoglu G, Tribbia J, Teng H (2012) A decadal prediction case study: late twentieth-century North Atlantic Ocean heat content. J Climate 25**:**5173-5189

Yu Y, Zheng W, Wang B, Liu H, Liu J (2011) Versions g1.0 and g1.1 of the LASG/IAP Flexible Global Ocean-Atmosphere-Land System model. Advances in Atmospheric Sciences 28**:**99-117

Yukimoto S, Adachi Y, Hosaka M, Sakami T, Yoshimura H, Hirabara M, Tanaka TY, Shindo E, Tsujino H, Deushi M, Mizuta R, Yabu S, Obata A, Nakano H, Koshiro T, Ose T, Kitoh A (2012) A new global climate model of the Meteorological Research Institute: MRI-CGCM3 - Model Description and Basic Performance -. Journal of the Meteorological Society of Japan 90A**:**23-64

Yukimoto S, Noda A, Kitoh A, Hosaka M, Yoshimura H, Uchiyama T, Shibata K, Arakawa O, Kusunoki S (2006) Present-day climate and climate sensitivity in the Meteorological Research Institute coupled GCM Version 2.3 (MRI-CGCM2.3). Journal of the Meteorological Society of Japan Ser II 84**:**333-363

Zanchettin D, Bothe O, Müller W, Bader J, Jungclaus J (2014) Different flavors of the Atlantic multidecadal variability. Climate Dyn 42**:**381-399

Zhang L, Wang C (2013) Multidecadal North Atlantic sea surface temperature and Atlantic meridional overturning circulation variability in CMIP5 historical simulations. J Geophys Res 118**:**5772–5791

Zheng F, Li J, Clark RT, Nnamchi HC (2013) Simulation and projection of the Southern Hemisphere Annular Mode in CMIP5 models. J Climate 26**:**9860-9879