LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

UCRL-JRNL-211903

# ENSO Simulation in Coupled Ocean-Atmosphere Models: Are the Current Models Better?

K. AchutaRao, K. R. Sperber

May 2, 2005

Climate Dynamics

# ENSO Simulation in Coupled Ocean-Atmosphere Models: Are the Current Models Better?

Krishna AchutaRao and Kenneth R. Sperber

Program for Climate Model Diagnosis and Intercomparison
Lawrence Livermore National Laboratory
P.O. Box 808, L-103
Livermore, CA 94550 USA

Climate Dynamics
December 2005
(revised)

Corresponding Author:
Dr. Krishna AchutaRao
Program for Climate Model Diagnosis and Intercomparison
Lawrence Livermore National Laboratory
P.O. Box 808, L-103
Livermore, CA 94550 USA
Phone: (925) 422-4197
Fax: (925) 422-7675
E-mail: achutarao1@llnl.gov

**Abstract** Maintaining a multi-model database over a generation or more of model development provides an important framework for assessing model improvement. Using control integrations, we compare the simulation of the El Niño/Southern Oscillation (ENSO), and its extratropical impact, in models developed for the 2007 Intergovernmental Panel on Climate Change (IPCC) Fourth Assessment Report with models developed in the late 1990's (the so-called Coupled Model Intercomparison Project-2 [CMIP2] models). The IPCC models tend to be more realistic in representing the frequency with which ENSO occurs, and they are better at locating enhanced temperature variability over the eastern Pacific Ocean. When compared with reanalyses, the IPCC models have larger pattern correlations of tropical surface air temperature than do the CMIP2 models during the boreal winter peak phase of El Niño. However, for sea-level pressure and precipitation rate anomalies, a clear separation in performance between the two vintages of models is not as apparent. The strongest improvement occurs for the modelling groups whose CMIP2 model tended to have the lowest pattern correlations with observations. This has been checked by subsampling the multi-century IPCC simulations in a manner to be consistent with the single 80-year time segment available from CMIP2. Our results suggest that multi-century integrations may be required to statistically assess model improvement of ENSO. The quality of the El Niño precipitation composite is directly related to the fidelity of the boreal winter precipitation climatology, highlighting the importance of reducing systematic model error. Over North America distinct improvement of El Niño forced boreal winter surface air temperature, sea-level pressure, and precipitation rate anomalies in the IPCC models occurs. This improvement, is directly proportional to the skill of the tropical El Niño forced precipitation anomalies.

## 1 Introduction

In advance of the Intergovernmental Panel on Climate Change Fourth Assessment Report (IPCC AR4) climate modelling groups from around the world have performed an unprecedented suite of simulations in order to assess the possible anthropogenic effects of greenhouse gases on the Earth's climate. An integral component of the IPCC AR4 is an evaluation of the mean state and the variability of the models under pre-industrial and present-day conditions to understand the strengths and weaknesses of the models. This will provide a baseline against which the climate change simulations can be compared.

Using control simulations, our goal is to compare and contrast the ability of the IPCC models to simulate the El Niño/Southern Oscillation (ENSO) with respect to the previous generation of models developed in the mid- to late 1990's. Since Southern Oscillation phenomena were first identified in a coupled ocean-atmosphere model (Sperber et al. (1987), numerous intercomparisons have been made of simulated ENSO variability. Neelin et al. (1992) found that numerous coupled feedback mechanisms for interannual variability were present in the early models, and that there was a "lack of robustness" in the ability of the models to simulate the "extent and position of the equatorial cold tongue and the western Pacific warm pool." Climate drift was also noted as major problem since, in some cases, it occurred via coupled processes that were not entirely distinct from those present in ENSO, and because of the unrealistic time-mean state to which many of the models evolved. Latif et al. (2001) found that few models were able to simulate the "gross equatorial SST (sea surface temperature) anomalies realistically" with the models overestimating (underestimating) SST variability in the western (eastern) tropical Pacific Ocean. Using many of the same models, Davey et al. (2002) linked the weak SST variability to errors in the wind stress, and noted that "most models have difficulty in reproducing the observed Pacific 'horseshoe' pattern of negative SST correlations with interannual Niño3 SST anomalies..." AchutaRao and Sperber (2002) placed an increased emphasis on assessing the ability of the models to simulate the spatial structures of the Walker circulation anomalies, and the warming and en-

hanced rainfall over the central/eastern tropical Pacific. More than two-thirds of the CMIP2 models they analyzed displaced these key ENSO signatures westward of their observed locations with the anomalies tending to be much weaker than observed, and many of the models exhibited a biennial peak in the power spectra of NINO3 temperature. This paper is an extension of the analysis of AchutaRao and Sperber (2002; hereafter referred to as AS02) using the more recent IPCC models. Applying this same methodology allows us to directly compare the fidelity with which ENSO is simulated in the two sets of models. Additionally, we have performed a more detailed analysis of ENSO teleconnections over the North American region, and we have included a statistical analysis to link this teleconnection to the quality of the El Niño forced precipitation in the tropics.

Numerous other evaluations of ENSO in the IPCC models have been performed. Wittenberg et al. (2005) analyzed the two IPCC GFDL models finding that they have a realistic ENSO period, including interdecadal variability. They noted the maximum ENSO SST anomalies were too strong as a result of being "too weakly damped by the surface heat fluxes", and they were displaced west of the observational maximum. This in turn affected the ENSO related teleconnection in the vicinity of North America. With a subset of the IPCC models Capotondi et al. (2005) note the applicability of the recharge oscillator of thermocline variations for ENSO variability, while Guilyardi (2005) and van Oldenborgh et al. (2005) also note the presence of an SST mode using a more complete suite of IPCC models. In the IPCC climate change simulations Merryfield (2005, Journal of Climate, submitted) and van Oldenborgh et al. (2005) find statistically significant changes in ENSO amplitude when the climate change signal is interpreted in the context of a model's internal variability from the control integration. However, there is no consensus on the sign of the amplitude change, and the changes are smaller than observed interdecadal variations. Merryfield (2005) suggests that the sign of the ENSO amplitude change is related to how broad the tropical Pacific ENSO windstress signal is and whether a model is dominated by an SST mode or a thermocline mode of ENSO variability. In the CMIP2 climate change simulations Collins et al. (2005) found no trend in ENSO, though they could not exclude

a switch to El Niño-like conditions with a small probability. The complexity of the feedbacks and interactions that drive and affect ENSO and the quality of the simulation of ENSO, the seasonal cycle, and the time-mean state in the control/present-day conditions requires continuing close scrutiny due to the important ramifications of forecasting ENSO on interannual and longer time scales.

In Section 2 we give brief details of the models and validation data used in this study. In section 3 we assess improvements in the simulation of ENSO, including extratropical teleconnections over the Northern Hemisphere, and our conclusions and further discussion are contained in Section 4.

## 2 Models and Validation Data

### 2.1 The models

Table 1 contains basic information on the IPCC models studied in this paper and their configurations. It also lists the number of years of model simulation analyzed, which ranges from 100-500 years in duration. (Note: the CMIP2 models analyzed by AS02 were all of 80 years duration). These are control simulations in which greenhouse gases are held fixed at preindustrial levels. Numerous modelling groups submitted data from more than one model version. The two GFDL models differ in their dynamical core, cloud scheme, and land model. The atmosphere and ocean component models in GISS-AOM differ from those of GISS-EH and GISS-ER. These latter two models only differ in the choice of ocean model. The two MIROC models employ the same physics, but are configured at different horizontal and vertical resolutions. From the National Center for Atmospheric Research (NCAR) results were submitted from the Parallel Climate Model (PCM), which was also used in CMIP2, and from the Community Climate System Model Version 3 (CCSM3). The UK Met Office HadCM3 IPCC data is from a later portion of the same integration used in CMIP2, and they also contributed data from their latest coupled model, HadGEM1. Auxiliary information about the IPCC models can be found at:
http://www-pcmdi.llnl.gov/ipcc/model_documentation/ipcc_model_documentation.php
The configurations of the CMIP2 models are discussed in AS02 with additional information at:

http://www-pcmdi.llnl.gov/projects/modeldoc/cmip/index.html

In order to perform a consistent analysis between the IPCC and the CMIP2 models we also analyze the surface air temperature since SST was unavailable from the CMIP2 simulations (AS02). We did not find any significant difference in our analysis of IPCC models by using the surface air temperature instead of the SST. Of the 19 IPCC models analyzed herein, 15 have previous generation counterparts that were evaluated by AS02, and this subset of models can be used to directly evaluate the impact of model development.

2.2 Validation data

Observed precipitation estimates from the Global Precipitation Climatology Project (GPCP; Huffman 1995, 1997) are used for model validation. GPCP is a merged product that incorporates satellite microwave and infrared data with rain gauge observations and covers the period (1979-2000). For surface air temperature and sea-level pressure reanalysis from the European Centre for Medium-Range Weather Forecasts Reanalysis-40 (ERA40) is the primary validation dataset. Additionally, data from National Centers for Environmental Prediction/National Center for Atmospheric Research (NCEP/NCAR) reanalysis is presented where appropriate. Multiple sources of "observational" data provide a rough estimate of observational uncertainty within which to assess model performance. ERA40 is a second generation reanalysis that uses the integrated forecast system that was operational from June 2001 to January 2002. The model was run at a horizontal resolution of T159 with 60 vertical levels. Further details of ERA40 can be found in Kallberg et al. (2005) and Uppala et al. (2005). The NCEP/NCAR Reanalysis was a joint project between NCEP and NCAR to produce a multi-decadal record of global atmospheric analyses with a fixed data assimilation system (Kalnay et al. 1996). This version of the model was operational in January 1995, and was run at a horizontal resolution of T62 with 28 vertical levels. From NCEP/NCAR reanalysis we analyze the period 1949-1998, and in limited cases 1979-93 to be consistent with ERA15 (Gibson et al. 1997; the first generation reanalysis from ECMWF). The

Hadley Centre Ice and Sea Surface Temperature (HadISST; Rayner et al. 2003) version 1.1 dataset is used to demonstrate the consistent spectral features and regional variations between SST and those derived from the reanalysis surface air temperature.

## 3 The El Niño/Southern Oscillation

In this section we will examine how well the models represent the temperature variability over the tropical Pacific in terms of amplitude, location, and power spectra. We also examine El Niño composites of surface air temperature, sea-level pressure, and precipitation for the boreal winter season to gauge improvement in an IPCC model compared to its CMIP2 predecessor. We will examine the relationship of the ENSO performance to the quality of the mean state and quantify the skill of El Niño forced teleconnection over North America, and relate it to the quality of the tropical forcing.

### 3.1 ENSO indices

Using monthly data AS02 evaluated the variability of surface air temperature over the NINO3 (5$^{o}$N-5$^{o}$S, 150$^{o}$W-90$^{o}$W), NINO3.4 (5$^{o}$N-5$^{o}$S, 170$^{o}$W-120$^{o}$W), and NINO4 (5$^{o}$N-5$^{o}$S, 160$^{o}$E-150$^{o}$W) regions. As seen in Fig. 1, reanalyses indicate that the standard deviation of air temperature increases from the western Pacific to the eastern Pacific Ocean, consistent with the results for SST using the HadISST data. For the models we only plot data from the NINO index that exhibited the largest standard deviation. Thus, we are ascertaining the ability of the models to concentrate their largest temperature variations in the NINO3 region, as observed. In CMIP2 numerous models incorrectly produced peak variability over the NINO4 region with only 7/17 (41%) models correctly simulating maximum variability over the NINO3 region. Of the IPCC models, only 1/19 exhibits the largest variation in the NINO4 region, while the majority of models (11/19; 58%) correctly represent the maximum variability in the NINO3 region. (For the IPCC models the numerical values of standard deviations for each of the regions are given in Table 2. These can be directly compared with those of the CMIP2 models in

Table 2 of AS02). Relative to their CMIP2 counterparts, 8/17 IPCC models were more realistic by shifting their maximum temperature variability further east, while the remaining 9 showed no change in the region of peak variability. The CMIP2 models, with two exceptions, tended to underestimate the standard deviation of the temperature in all three regions, whereas the IPCC models display a broader range of variability with many models overestimating the temperature variability. For the two IPCC GFDL models (E1 and E2) there is a 32% difference in their standard deviations of the NINO3.4 temperature, suggesting a substantial impact due to the different dynamical cores, cloud schemes, and land models. All of the IPCC GISS models exhibit peak surface air temperature variability over NINO3, but F1 and F2, which differ in both atmosphere and ocean component models, greatly underestimate the observed standard deviation similar to their CMIP2 entry, f. However, F3, which has the same atmospheric model as F2 but a different ocean model, has a realistic level of variability indicating pronounced sensitivity to the choice of ocean model. DOE-PCM was used in IPCC and CMIP2 (M and m) and they are robust in their standard deviation estimate of NINO3 temperature. However, for HadCM3 the CMIP2 contribution has 23% greater variability than its HadCM3 counterpart. In order to eliminate the possibility of the longer time-series in the IPCC submission causing the lowered standard deviation, we calculated the NINO3 standard deviation over individual 80 year segments of the HadCM3 IPCC contribution that overlap by all but 10 years. We found that the standard deviations range between 0.74-1.00, still lower than the CMIP2 estimate. This difference in variability between the CMIP2 and IPCC contributions suggests that there is interdecadal ENSO variability in this earlier segment of the integration that was not represented in the (distinct) segment analyzed in IPCC. The HadCM3 IPCC (CMIP2) contribution has the largest variability over NINO3 (NINO3.4) though the difference in the standard deviations between the two regions is only 0.01K in each simulation, respectively. AS02 showed that the standard deviation of NINO3 surface air temperature from NCEP/NCAR reanalysis ranged from approximately 0.5 to 0.85$^{o}$C on interdecadal time scales using a 15-year moving window.

For the NINO3 region, we show in Fig. 2 power spectra of surface air temperature from the IPCC and the CMIP2 models. Relative to the reanalyses and the HadISST SST the IPCC models tend to have a more a realistic ENSO time scale compared to the CMIP2 models. Approximately 25% of the IPCC models fail to simulate a spectral peak for periods of 2 to 7 years (Fig. 2a), and these models are those with weak tropical SST variability. This is an improvement compared to CMIP2 in which nearly 50% of the models exhibited this shortcoming (Fig. 2b). Also, fewer IPCC models exhibit dominant power at biennial time scales. Several IPCC models have greater power than observed (see also Table 2).

### 3.2 Composites of ENSO Events

Using ERA40 reanalyses and GPCP data global composite anomalies of surface air temperature, sea-level pressure and precipitation have been generated for December-February (DJF) for the peak of the warm and cold phases of ENSO. As in AS02 we require a consistent atmosphere-ocean response to clearly depict the El Niño (La Niña) events such that the standardized DJF NINO3 surface air temperature anomalies be greater than or equal to 0.6 (less than or equal to -0.6) and the standardized DJF SOI be less than or equal to -0.6 (greater than or equal to 0.6). These thresholds, derived from observations, allow us to recover the observed events over the period 1949-1998. For constructing the spatial composites of observed precipitation we only use the ENSO events contained in the shorter period of the GPCP dataset (1979-2000). The observed composites for El Niño, given in Figs. 3a-c, depict the tropical Pacific warming that extends from the dateline to the coast of South America, the perturbation to the Walker circulation, and the enhancement of tropical rainfall over the central/eastern Pacific Ocean. Improvement in the representation of these key spatial patterns of El Niño anomalies are shown using composites of the CMIP2 and IPCC runs from the Meteorological Research Institute (MRI) models in Figs. 3d-3f and 3h-3i. The MRI CMIP2 model severely underestimates the surface air temperature anomalies over the eastern Pacific, the positive sea-level pressure anomalies over the western Pacific are too weak, and the strongest positive rainfall anomalies are displaced at or west of the date-

line. The MRI IPCC model gives an excellent representation of the observed El Niño composite patterns over the tropical Pacific, demonstrating increased skill in representing ENSO through model development. It is not possible to isolate the causal factors that resulted in the improved El Niño composites due to the whole scale modifications to the coupled model (e.g, a new atmospheric GCM).

To assess model performance we utilize the Taylor plot (Taylor 2001). In this application the model composite is compared to an "observed" reference composite with the spatial standard deviation and its associated root-mean-square-difference (RMSD) plotted in relation to the pattern correlation. These metrics for El Niño composites of surface air temperature over the tropical domain ($30^{o}$N-$30^{o}$S, $90^{o}$E-$90^{o}$W; see the black box in Fig. 3a) are shown in Fig. 4. Relative to ERA40 the result from NCEP/NCAR reanalysis is also plotted, indicating that the models lie outside of the range of observational uncertainty. Compared to ERA40 the majority of CMIP2 models underestimated the spatial standard deviation, whereas the IPCC models span a broader range of values. The majority of IPCC models have larger pattern correlations and smaller RMSDs (when compared with ERA40) than the CMIP2 models. Errors in the RMSD arise for two reasons: (1) a poor match between the simulated and observed spatial patterns, and (2) errors in the magnitude of the anomalies. That the IPCC models tend to have larger pattern correlations than the CMIP2 models indicates that errors in the magnitude of the anomalies remain a substantial component of the RMSD (see Fig. 1 and Table 2). Taylor plots for sea-level pressure and precipitation (not shown) suggest that there is no clear separation of performance between the two vintages of models.

An important difference between CMIP2 and IPCC is the length of the control simulation that is archived. In CMIP2 only 80 years of monthly mean data were archived (AS02) whereas in the IPCC database longer simulations are archived (Table 1). As seen with the differences in variability in the NINO3 region of the CMIP2 and IPCC runs of the HadCM3 model, there can be differences within the same run because of the length and segment of the simulation analyzed. In forming composites, longer simulations will mean that a larger number of ENSO events are used to generate the composite

pattern which in turn may result in a smoother composite that can give rise to larger pattern correlations with observations. The smoother composite patterns would be due to a reduction in random heterogeneity that occurs during individual events, as well as sampling a larger variety of the "flavors of El Niño" including decadal modulation in the strength of El Niño. In order to account for differences resulting from the analysis of unequal lengths of simulations, we analyzed contiguous 80-year segments from the IPCC runs that overlap by all but 10 years (e.g., years 1-80, 11-90, 21-100, etc.). For each 80 year segment we recompute the El Niño composite anomalies and their pattern correlation relative to observations. While the segments are not completely independent, they do provide an estimate of the distribution of composite statistics. The spread of an IPCC model's pattern correlations in the Tropical region is shown as a bar and whisker plot against the pattern correlation for that model's CMIP2 counterpart in Fig. 5. The ends of the whiskers indicate the smallest and largest IPCC composite pattern correlations, the lower and upper limits of the box correspond to the 25th and 75th percentiles, and the dividing line within the box is the median pattern correlation. The width of the boxes is proportional to the number of 80-year samples derived from the model's IPCC control run. For each model a filled circle indicates the pattern correlation for the composite derived from the full IPCC run. A diagonal line indicating equal pattern correlations in the IPCC and CMIP simulations is shown for visual reference.

These plots reveal whether the range of pattern correlations of the composite anomalies in the IPCC simulation are different from those of the CMIP2 simulation. We consider the IPCC simulation to be different from the CMIP2 simulation if the box and whisker do not cross the diagonal. Where the box and whisker lie completely above (below) the diagonal, the IPCC simulation has an improved (degraded) pattern correlation compared to its CMIP2 counterpart. Where the box and whisker intersect the diagonal, there is no clear difference between the IPCC and CMIP2 simulations. With the exception of M and N1 (see Table 1 for model identification) we are assessing the effect of model development from the individual modelling groups. For these two models, this plot reveals differences

between the segments analyzed. We first discuss the 15 IPCC models that are different model versions from their CMIP2 counterparts.

In the case of surface air temperature (Fig. 5a), 9/15 IPCC models (60%) always have higher pattern correlations with ERA40 than their predecessor CMIP2 model. (i.e., the distribution of pattern correlations given by their bar and whisker plots all lie above the diagonal). It is noteworthy that most of the CMIP2 models that had pattern correlations <0.6 are among the models that have shown improvement. One model (N2) has degraded compared to its CMIP2 predecessor. The remaining 5 models (F1, F2, B1, B2 and D) have the box and whisker straddling the diagonal indicating that there is no change in performance. The IPCC models display diverse behavior in the spread of pattern correlations they produce even in cases where the number of 80 year segments analyzed is the same (B1 vs. E1 or E2). This indicates that the spread of correlations appears to be dependent on the model and less so on the number of 80-year samples (indicated by the width of the box) analyzed. Given the short length of the observational record we do not know the form of the observed distribution of pattern correlations, and hence what spread in the distribution is actually realistic.

The pattern correlations derived from the full length of the IPCC data (indicated by the filled circles) tend to be at the upper end of the whisker and in some cases (e.g., K, N1, and N2), being greater than that for any individual 80 year segment - possibly as a result of smoothing over a large number of individual events. A thorough investigation of the reasons for this are beyond the scope of this paper. An alternative interpretation is that using a longer record length provides a more robust estimate of a models ability to simulate the El Niño composites. Under this assumption 11/15 (73%) of IPCC models have improved pattern correlations of surface air temperature compared to their predecessor CMIP2 model. The discrepancy occurs for models F2 and D whose bar and whisker plots for the individual 80 year segments have some pattern correlations that lie below the diagonal, while for the full record the pattern correlation lies above the diagonal. It therefore becomes useful to consider equal length samples from the CMIP2 and IPCC models. Other than the fact that the CMIP2 data were available for 80

years there was no *a priori* reason for subsampling the IPCC data in 80 year segments. The diversity of the distributions of the IPCC pattern correlations from analyzing 80 year segments and the off-diagonal distributions of precipitation from M and N1 suggest that multi-century simulations are necessary in order to assess the ability of a model to simulate ENSO and gauge its improvement relative to previous model versions.

As with surface air temperature, the tropical pattern correlations of El Niño composite anomalies of sea-level pressure and precipitation show the most improvement for groups whose CMIP2 models tended to have low pattern correlations with ERA40 and GPCP precipitation (Figs. 5b and 5c, respectively). However, the pronounced improvement noted for surface air temperature is not evident for sea-level pressure and precipitation composites. For sea-level pressure (precipitation) only 5 (4) IPCC models shows improved composites compared to CMIP2, while 4 (9) IPCC models show degraded performance compared to CMIP2 (based on the IPCC 80-year subsampled pattern correlation distributions not intersecting the diagonal in Fig. 5). As in the case of surface air temperature, all of the improvements have been realized in models whose CMIP2 model composite had low pattern correlation. For both surface air temperature and sea-level pressure the majority of IPCC models have pattern correlations with ERA40 that exceed 0.6. However, a noticeable degradation in the IPCC models occurs for precipitation composites (Fig. 5c). Whereas many of the CMIP2 models had pattern correlations >0.6, their IPCC counterparts exhibit lower pattern correlations. Given the wide scope of model development, it is not possible to isolate what aspect of the modified physics or dynamics has caused this degradation. We provide some possible reasons for this degradation in performance in the discussion section. Overall, only 3 IPCC models (G, H, and J - the MRI-CGCM2.3.2 presented in Fig. 3) show improvement in pattern correlations for all three variables based on our methodology of using 80-year long samples. Suggested improvement for all three variables extends to 5 additional IPCC models (A, E1, E2, F2, and K) if one considers the full record length. We performed a similar analysis (not shown) using the RMSD of the model composites with respect to the composites from observations. Our con-

clusions based on the increases (decreases) in pattern correlations are confirmed by corresponding decreases (increases) in RMSD.

Numerous studies have indicated that the realism of variability is directly related to an accurate simulation of the mean state. In simulations forced with observed SST Sperber and Palmer (1996) and Sperber et al. (1999) showed that models that had a more realistic boreal summer precipitation climatology over India were better able to capture the monsoon-ENSO teleconnection. Using HadCM3, Turner et al. (2005) showed that improving the mean state of the tropical Pacific through the use of flux-adjustment gave rise to a more realistic ENSO that in turn improved the summer monsoon-ENSO teleconnection. Annamalai et al. (2005) find that realistic representation of the summer monsoon-ENSO relationship is predicated on simulating realistic time mean rainfall over the monsoon region and properly representing the ENSO evolution of sea surface temperature and diabatic heating in the tropical Pacific. AS02 had noted that the CMIP2 models with more realistic composite patterns also had more realistic precipitation climatologies for the DJF season. In Fig. 6 we show the relationship between the pattern correlations of the precipitation rate climatology and the El Niño composite in the tropical Pacific. Our results indicate that the quality of the El Niño precipitation composite is directly proportional to the quality of the time-mean rainfall during the boreal winter season. The relationship is statistically significant at the 5% level using a two-tailed t-test for both the CMIP2 and IPCC models. This further highlights the importance of reducing systematic error in models.

### 3.3 The Extratropical Response to El Nino

Perturbations of sea-level pressure and surface air temperature in the vicinity of North America were first linked to the Southern Oscillation by Walker and Bliss (1930), and have been investigated in increasing detail over the past 40 years. Observational studies by Bjerknes (1966, 1969), Horel and Wallace (1981), van Loon and Madden (1981), Barnston and Livezey (1987), and Trenberth and Caron (2000) for example, established that the teleconnections tend to be strongest during boreal winter when

the ENSO forced diabatic heating anomalies in the central Pacific are strongest. While many papers have shown that general circulation models simulate well the geopotential height anomalies associated with ENSO (e.g., Kumar and Hoerling 1997), extratropical signals in sea-level pressure, surface temperature, and precipitation also need to be considered. For surface air temperature, the extratropical response is characterized by warming over Canada, extending to the Great Lakes and into the northeastern United States (Fig. 3a). Negative sea-level pressure anomalies prevail over much of North America, especially over the Pacific northwest and off the east coast of the United States, whereas above normal sea-level pressure anomalies occur over northeastern Canada (Fig. 3b). With the reduced sea-level pressure off the west coast of the North America the westerly jet tends to be more zonal, with the result that rainfall is more plentiful near the west coast of the United States (Fig. 3c). Enhanced rainfall along the Gulf coast and the southeastern United States are also common features observed during El Niño. In this section we compare and contrast the ability of the CMIP2 and the IPCC models to capture these aspects of the extratropical response to El Niño over the North American region ($20^{o}$N-$70^{o}$N, $140^{o}$W-$60^{o}$W; see the red box in Fig. 3a).

Using Taylor plots we assess the ability of the models to capture the December-February El Niño composites over the "North American" region shown in Fig. 3 by the red box. The results in Fig. 7a are for surface air temperature anomalies as compared to ERA40. The NCEP/NCAR reanalysis is shown along with the models to give a measure of how far the models are from a second observationally based dataset. The IPCC models tend to perform better than the CMIP2 models with the majority having pattern correlations (RMSDs) in excess (below) of 0.2 (1.0). Similarly, for sea-level pressure and precipitation the IPCC models exhibit improved Taylor statistics compared to the CMIP2 models (Figs. 7b and 7c). The models best represent the sea-level pressure composite, followed by precipitation and surface air temperature. This suggests that the precipitation signal may be more predictable than the temperature signal in the IPCC models, consistent with the analysis of a subset of IPCC models using the 20th century integrations (Joseph and Nigam 2005, Journal of Climate, submitted). As

with the tropics, subsetting the IPCC model data into 80 year segments to create a distribution of pattern correlations for the North America region indicates that the improvement tends to be strongest for groups whose CMIP2 models were most problematic (not shown).

AS02 suggested that a westward displacement of diabatic heating anomalies in the tropical Pacific could give rise to a westward displacement of the warm anomalies that typically occur over the northern tier of North America (Fig. 3a). Hoerling and Kumar (2002) performed sensitivity experiments using an atmospheric GCM in which idealized warm SST anomalies were prescribed at different locations in the tropical Pacific. A shift of the associated diabatic heating from 170$^{o}$E to 140$^{o}$W caused a reversal of the sign of the precipitation anomalies along the west coast of North America. In the GFDL-CM2.0 and GFDL-CM2.1 models Wittenberg et al. (2005) noted that the westward displacement of the ENSO related 200hPa geopotential heights over Canada is associated with a similar displacement of the tropical ENSO forcing. These results establish that simulating the correct location of the maximum diabatic heating in the tropics is essential in order to correctly represent the observed teleconnection between El Niño and the extratropics. Given the importance of the ENSO forced tropical diabatic heating for establishing the extratropical teleconnection, in Fig. 8 we show the pattern correlations of the DJF El Niño composites over North America in relation to the concomitant tropical precipitation composite pattern correlations. The results are shown for the individual 80 year segments (colored dots) and the full record (lettered circles) of the IPCC simulations. For all three variables linear regression indicates that a better representation of the North American teleconnection is associated with a better representation of the tropical El Niño precipitation composite pattern correlations. This result, based on the 80 year segments, is statistically significant at the 1% level using a two-tailed t-test that assumes no *a priori* relationship. A closer look at precipitation, Fig. 8a, shows that some models have robust pattern correlations in the tropics and over North America with little spread indicated by the 80 year segments (e.g., N2, V, C, N1, and J). Thus it is clear that across models a higher pattern correlation over the tropical Pacific yields a better pattern correlation over North America. Proportion-

ate skill between the tropics and North America also holds for individual models that exhibit a wide range of skill in their 80 year segments (e.g., M, H, U, and A). The close association between "tropical" skill and the "North American" skill is readily apparent for MRI-CGCM2.3.2 as seen in Figs. 3a-3c and 3g-3i. It should be kept in mind that there may be other factors contributing to the improved El Niño teleconnection pattern over North America, such as an improved representation of the extratropical longwave pattern that would facilitate the tropical-extratropical linkage (e.g, Joseph and Nigam 2005).

**4 Summary and Discussion**

The analysis of ENSO simulations in the CMIP2 models by AS02 is used as a benchmark for evaluating the latest IPCC models developed by many of the worlds climate modelling centers. Relative to their CMIP2 counterparts, nearly two-thirds of the IPCC models exhibited increased pattern correlations of surface air temperature and sea-level pressure anomalies over the tropics during the boreal winter peak phase of El Niño. However, most IPCC models show decreased pattern correlation of precipitation rate anomalies relative to their CMIP2 counterparts. The positive rainfall anomalies tend to extend too far west along the equator, and this error is more pervasive than in the CMIP2 models. This may be associated with the tendency for the IPCC models to produce a split ITCZ over the western Pacific. A contributing factor may be the relative lack of the use of flux-adjustment in IPCC models (2) compared to the CMIP2 models (10) since this artificial constraint helps maintain a realistic basic state. During the boreal winter peak phase of El Niño all of the IPCC models that improved their composite tropical precipitation rate anomaly pattern correlations also improved their pattern correlations of surface air temperature and sea-level pressure anomaly composites in the tropics. Importantly, the quality of the El Niño composite precipitation rate anomalies is directly proportional to the quality of the boreal winter tropical precipitation rate climatology, thus underscoring the importance of reducing errors in the time-mean state. Improvement in the composite tropical precipitation rate anomaly pattern

correlation also gives rise to a statistically significant improvement of pattern correlations of surface air temperature, sea-level pressure, and precipitation rate composites over the North American region. This demonstrates the sensitivity of this key El Niño teleconnection pattern to the fidelity with which the tropics is represented.

Over the past 20 years dramatic improvement in our ability to simulate ENSO has occurred. Some of this improvement can be attributed to improved physical parameterizations in the ocean, such as vertical mixing and the computing power necessary to run the ocean models at high enough resolution to begin to resolve equatorial wave dynamics. These aspects of coupled model formulation probably account for a substantial fraction of the improvement in ENSO simulation that occurred between the Neelin et al. (1992) and the ASO2 intercomparisons. However, these improvements are not sufficient conditions for a realistic ENSO simulation. The sensitivity of the east-west temperature gradient and the variabilty of tropical Pacific SST due to physics changes within an atmospheric model have been noted by Schneider (2002) and due to changes in an ocean model by Meehl et al, (2001). Guilyardi et al. (2004) used different model configurations that shared either the same ocean or atmospheric components. They found that the atmospheric model played the dominant role in setting the periodicity of ENSO, and that increased horizontal resolution in the atmospheric model gave a more realistic broadband spectrum of interannual variability as opposed to "a single, preferred-too high-frequency" in the lower resolution version. Thus, due to the coupled ocean-atmosphere nature of ENSO, constant re-evaluation of ENSO simulations in models is needed. The results presented in this study indicate that climate models tend to be more realistic than ever in their simulation of tropical ENSO variability, and its forced extratropical response in the vicinity of North America during boreal winter. This indicates that compared to the climate models used in the IPCC Third Assessment Report (Houghton et al. 2001), we should have increased confidence in predictions of the impact of anthropogenic forcing on this mode of variability for the IPCC models analyzed herein, though this is conditional since the models must also properly represent climate change feedback processes. These results underscore the sen-

sitivity of tropical-extratropical linkages, and the high standards to which model development should endeavor.

**Acknowledgments** We thank Drs. Ben Santer, Karl Taylor, and Jim Boyle for helpful conversations regarding sampling issues. We wish to thank Matt Collins and one anonymous reviewer for helpful comments and suggestions. We acknowledge the international modeling groups for providing their data for analysis, the Program for Climate Model Diagnosis and Intercomparison (PCMDI) for collecting and archiving the model data, the JSC/CLIVAR Working Group on Coupled Modelling (WGCM) and their Coupled Model Intercomparison Project (CMIP) and Climate Simulation Panel for organizing the model data analysis activity, and the IPCC WG1 TSU for technical support. The IPCC Data Archive at Lawrence Livermore National Laboratory is supported by the Office of Science, U.S. Department of Energy. This work was performed under the auspices of the U.S. Department of Energy Office of Science, Climate Change Prediction Program by University of California Lawrence Livermore National Laboratory under contract No. W-7405-Eng-48.

**References**

AchutaRao K, Sperber K (2002) Simulation of the El Niño Southern Oscillation: Results from the Coupled Model Intercomparison Project. Clim Dynam 19: 191-209

Annamalai H, Hamilton K, Sperber KR, (2005) South Asian summer monsoon and its relationship with ENSO in the IPCC AR4 simulations. J Clim (submitted)

Barnston A, Livezey RE (1987) Classification, seasonality, and persistence of low-frequency circulation patterns. Mon Wea Rev 115: 1083-1126

Bjerknes J (1966) A possible response of the atmospheric Hadley circulation to equatorial anomalies of ocean temperature. Tellus 18: 820-829

Bjerknes J (1969) Atmospheric teleconnections from the equatorial Pacific. Mon Wea Rev 97: 163-172

Capotondi A, Wittenberg A, Masina S (2005) Spatial and temporal structure of ENSO in 20th century coupled simulations. Ocean Modeling (submitted)

Collins M, The CMIP modelling groups (2005) El Niño- or La Niña-like climate change. Clim Dynam 24: 89-104

Davey MK, and coauthors (2002) STOIC: a study of coupled model climatology and variability in tropical ocean regions. Clim Dynam 18: 403-420

Gibson, JK, Kallberg P, Uppala S, Hernandez A, Nomura A, Serrano E (1997) ERA Description. ECMWF Reanalysis Project Report Series 1, European Centre for medium-Range Weather Forecasts, Reading, UK, 66pp

Guilyardi E (2005) El Niño - mean state - seasonal cycle interactions in a multi-model ensemble. Clim Dynam (in press)

Guilyardi E, Gualdi S, Slingo J, Navarra A, Delecluse P, Cole J, Madec G, Roberts M, Latif M, Terray L (2004) Representing El Niño in coupled ocean-atmosphere GCMs: The dominant role of the atmospheric component. J Clim 17: 4623-4629

Hoerling MP, Kumar A (2002) Atmospheric response patterns associated with tropical forcing. J Clim 15: 2184-2203.

Horel JD, Wallace JM (1981) Planetary scale atmospheric phenomena associated with the Southern Oscillation. Mon Wea Rev 109: 813-829

Huffman GJ, Adler, RF, Rudolf B, Schneider U (1995) Global precipitation estimates based on a technique for combining satellite-based estimates, rain gauge analysis, and NWP model precipitation information. J Clim 8: 1284-1295

Huffman GJ, Adler RF, Arkin P, Chang A, Ferraro R, Gruber A, Janowiak J, McNab A, Rudolf B, Schneifer U (1997) The global precipitation climatology project (GPCP) combined precipitation dataset. BAMS 78: 5-20

Houghton JT, Ding Y, Griggs DJ, Noguer M, van der Linden PJ, Dai X, Maskall K, Johnson CA (2001) Climate Change (2001) The Scientific Basis. Cambridge University Press, Cambridge, UK 881pp

Joseph R, Nigam S (2005) ENSO evolution and teleconnections in IPCC's 20th century climate simulations: realistic representation? J Clim (submitted)

Kallberg P, Berrisford P, Hoskins B, Simmons A, Uppala S, Lamy-Thepaut S, Hine R (2005) ERA40 Atlas. ECMWF Re-Analysis Project Report Series 19. ECMWF, Shinfield Park, Reading, UK 191pp

Kalnay E, Kanamitsu, M, Kistler R, Collins W, Deaven D, Gandin L, Iredell M, Saha S, White G, Woollen J, Zhu Y, Chelliah M, Ebisuzaki W, Higgins W, Janowiak J, Mo KC, Ropelewski C, Wang J, Leetma A, Reynolds R, Jenne R, Joseph D (1996) The NCEP/NCAR 40-year reanalysis project. Bull Amer Met Soc 77: 437-471

Kumar A, Hoerling MP (1997) Interpretation and implications of the observed inter-El Nino variability. J Clim 10: 83-91

Latif M, and coauthors (2001) ENSIP: the El Niño simulation intercomparison project. Clim Dynam 18: 255-276

Meehl GA, Gent PR, Arblaster JM, Otto-Bliesner BL, Brady EC, Craig A (2001) Factors that affect the amplitude of El Nino in global coupled climate models. Clim Dynam 17: 515-526

Merryfield WJ (2005) Changes in ENSO under CO2 doubling in the IPCC AR4 coupled climate models. J Clim (submitted)

Neelin JD, and coauthors (1992) Tropical air-sea interaction in general circulation models. Clim Dynam 7: 73-104

Rayner NA, Parker DE, Horton EB, Folland CK, Alexander LV, Rowell DP, Kent EC, Kaplan A (2003) Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. J Geophys Res, 108 (D14), 4407, doi:10.1029/2002JD002670

Schneider EK (2002) Understanding differences between the equatorial Pacific as simulated by two coupled GCMs. J Clim 15: 449-469

Sperber KR, Hameed S, Gates WL, Potter GL (1987) Southern oscillation simulated in a global climate model. Nature 329, 140-142

Sperber KR, Palmer TN (1996) Interannual tropical rainfall variability in general circulation model simulations associated with the Atmospheric Model Intercomparison Project. J Clim: 9 2727-2750

Sperber KR, and participating AMIP modelling groups (1999) Are revised models better? A skill score assessment regional interannual variability. Geophys Res Lett 26: 1267-1270

Taylor KE, (2001) Summarizing multiple aspects of model performance in a single diagram. J geophys Res 106: 7183-7192

Trenberth KE, Caron JM (2000) The Southern Oscillation revisited: sea level pressures, surface temperatures and precipitation. J Clim 13: 4358-4365

Turner AG, Inness PM, Slingo JM (2005) The role of the basic state in the ENSO-monsoon relationship and implications for predictability. Quart J Roy Meteor Soc 131**:** 781-804.

Uppala S, and coauthors (2005) The ERA-40 Re-Analysis. ECMWF Re-Analysis Project Report Series (in press)

van Loon H, Madden RA (1981) The Southern Oscillation. Part I: Global associations with pressure and temperature in the northern winter. Mon Wea Rev 109: 1150-1162

van Oldenborgh GJ, Philip S, Collins M (2005) El Niño in a changing climate: a multi-model study. Ocean Sci Discussions 2: 267-298

Walker GT, Bliss EW (1930) World Weather IV Some applications to seasonal foreshadowing. Mem Royal Meteorol Soc 3: 81-95

Wittenberg AT, Rosati A, Lau N-C, Ploshay JJ (2005) GFDL's CM2 global coupled climate models, Part 3: tropical Pacific climate and ENSO. J Clim (in press)

**Table Captions**

**Table 1**. Table of IPCC models analyzed in this study listed by modelling group. Information is also provided about the model designations, the horizontal and vertical resolution of the atmospheric and oceanic component models, and length in years of the pre-industrial control run analyzed. The corresponding lower case alphabetic character represents the predecessor CMIP2 model in the figures

**Table 2.** The standard deviations of monthly anomalies of temperature ($^{o}$C) over the NINO3 (5$^{o}$N-5$^{o}$S, 150$^{o}$W-90$^{o}$W), NINO3.4 (5$^{o}$N-5$^{o}$S, 170$^{o}$W-120$^{o}$W), and NINO4 (5$^{o}$N-5$^{o}$S, 160$^{o}$E-150$^{o}$W) regions. Numbers in bold represent the region with peak variability. For the reanalyses and the models the results are for surface air temperature. For HadISST the results are for sea surface temperature

**Table 1.** Table of IPCC models analyzed in this study listed by modelling group. Information is also provided about the model designations, the horizontal and vertical resolution of the atmospheric and oceanic component models, and length in years of the pre-industrial control run analyzed. The corresponding lower case alphabetic character represents the predecessor CMIP2 model in the figures

| Modelling Group[a] | Model Designation and Figure Symbol [ ] | AGCM Horizontal/ Vertical Resolution | OGCM Horizontal/ Vertical Resolution | Length of Run Archived |
|---|---|---|---|---|
| Canadian Centre for Climate Modelling & Analysis | CGCM3.1(T47) [A] | T47 L31 | 192x96 L29 | 500 |
| Center for Climate System Research (The University of Tokyo), National Institute for Environmental Studies, and Frontier Research Center for Global Change (JAMSTEC) | MIROC3.2(medres) [B1]<br>MIROC3.2(hires) [B2] | T42 L20<br>T106 L56 | 256x192 L44<br>T106 L48 | 500<br>100 |
| CSIRO Atmospheric Research | CSIRO-Mk3.0 [C] | T63 L18 | 1.875x0.925 L31 | 380 |
| Max Planck Institute for Meteorology | ECHAM5/MPI-OM[b] [D] | T63 L32 | 1.5x1.5 L40 | 332 |
| Geophysical Fluid Dynamics Laboratory (NOAA) | GFDL-CM2.0 [E1]<br>GFDL-CM2.1 [E2] | N45 L24<br>N45 L24 | 1x0.33-1 L50<br>1x0.33-1 L50 | 500<br>500 |
| Goddard Institute for Space Studies (NASA) | GISS-AOM [F1]<br>GISS-EH [F2]<br>GISS-ER [F3] | 90x60 L12<br>72x46 L17<br>72x46 L17 | 90x60 L16<br>2x2 cos(lat) L16<br>72x46 L13 | 251<br>400<br>500 |
| LASG/Institute of Atmospheric Physics | FGOALS-g1.0 [G] | 128x60 L26 | 360x170 L33 | 150 |
| Institut Pierre Simon Laplace | IPSL-CM4 [H] | 96x72 L19 | 2x2 L31 | 230 |
| Meteorological Research Institute | MRI-CGCM2.3.2 [J] | T42 L30 | 2x0.5-2.5 L23 | 350 |
| National Center for Atmospheric Research | CCSM3[c] [K]<br>PCM [M] | T85 L26<br>T42 L18 | <br>384x288 L32 | 230<br>350 |
| Hadley Centre for Climate Prediction and Research/Met Office | UKMO-HadCM3 [N1]<br>UKMO-HadGEM1 [N2] | 2.5x3.75<br>N96 L38 | 1.25x1.25 L20<br>1x0.33-1 L40 | 341<br>140 |
| Centre National de Recherches Météorologiques (Meteo-France) | CNRM-CM3 [U][d] | T42 L45 | 180x170 L33 | 390 |
| Institute for Numerical Mathematics | INM-CM3.0 [V][d] | 4x5 L21 | 2x2.5 L33 | 330 |

[a]Several CMIP2 models analyzed do not have IPCC counterparts or they are not used as the CMIP2 counterpart for for assessing the performance change of the IPCC model. Their model names (as in AS02) and figures symbols in this paper [ ] are: BMRC [p], CERFACS [q], ECHAM3+LSG [r], NCAR (WM) [s], and HadCM2 [t]

[b]Performance change is measured with respect to ECHAM4+OPYC3 from CMIP2

[c]Performance change is measured with respect to NCAR (CSM) from CMIP2

[d]These IPCC models do not have CMIP2 counterparts analyzed in AS02

**Table 2.** The standard deviations of monthly anomalies of temperature ($^{o}$C) over the NINO3 (5$^{o}$N-5$^{o}$S, 150$^{o}$W-90$^{o}$W), NINO3.4 (5$^{o}$N-5$^{o}$S, 170$^{o}$W-120$^{o}$W), and NINO4 (5$^{o}$N-5$^{o}$S, 160$^{o}$E-150$^{o}$W) regions. Numbers in bold represent the region with peak variability. For the reanalyses and the models the results are for surface air temperature. For HadISST the results are for sea surface temperature

| Model/Region | NINO4 | NINO3.4 | NINO3 |
|---|---|---|---|
| HadISST v1.1 | 0.55 | 0.76 | **0.80** |
| NCEP/NCAR | 0.48 | 0.66 | **0.74** |
| ERA15 | 0.51 | 0.87 | **0.95** |
| CGCM3.1(T47) | 0.41 | **0.42** | 0.38 |
| CNRM-CM3 | 1.32 | 1.63 | **1.80** |
| CSIRO-Mk3.0 | 0.90 | 0.98 | **1.00** |
| ECHAM5/MPI-OM | 1.15 | **1.36** | **1.36** |
| GFDL-CM2.0 | 0.84 | **0.93** | 0.86 |
| GFDL-CM2.1 | 1.10 | **1.23** | 1.20 |
| GISS-AOM | 0.12 | 0.12 | **0.13** |
| GISS-EH | 0.55 | 0.74 | **0.80** |
| GISS-ER | 0.10 | 0.12 | **0.15** |
| FGOALS-g1.0 | 1.55 | **1.83** | **1.83** |
| INM-CM3.0 | **0.98** | 0.90 | 0.94 |
| IPSL-CM4 | 0.61 | 0.89 | **0.92** |
| MIROC3.2(hires) | 0.27 | **0.29** | **0.29** |
| MIROC3.2(medres) | 0.45 | **0.48** | **0.48** |
| MRI-CGCM2.3.2 | 0.85 | **1.03** | 0.89 |
| CCSM3 | 0.60 | 0.92 | **0.95** |
| PCM | 0.62 | 0.81 | **0.84** |
| UKMO-HadCM3 | 0.69 | 0.85 | **0.86** |
| UKMO-HadGEM1 | 0.52 | **0.72** | 0.71 |

**Figure Captions**

**Fig. 1.** Indices of monthly anomalies of surface air temperature ($^{o}$C) have been calculated for the NINO3 (5$^{o}$N-5$^{o}$S, 150$^{o}$W-90$^{o}$W), NINO3.4 (5$^{o}$N-5$^{o}$S, 170$^{o}$W-120$^{o}$W), and NINO4 (5$^{o}$N-5$^{o}$S, 160$^{o}$E-150$^{o}$W) regions. For the models we only plot the standard deviation for the index with the largest standard deviation. Results from the CMIP2 (IPCC) models are given by the unshaded (shaded) circles. Estimates of the observed standard deviation for each index are given from numerous sources, indicating that the observed standard deviation is largest in NINO3 and smallest in NINO4. Note: The standard deviation from HadISST is for sea surface temperature. See Table 1 for model identification notation

**Fig. 2.** Maximum entropy power spectra of NINO3 (5$^{o}$N-5$^{o}$S, 150$^{o}$W-90$^{o}$W) surface air temperature for **a** the IPCC models and **b** the CMIP2 models (after Fig. 4 in AchutaRao and Sperber 2002). The vertical lines correspond to periods of 2 and 7 years. The power spectra from the reanalyses and for SST from the HadISST version 1.1 dataset are given by the series of solid, dashed, and dotted black curves

**Fig. 3.** December-February warm event composite anomalies of **a** surface air temperature ($^{o}$C), **b** sea-level pressure (hPa), **c** precipitation rate (mm day$^{-1}$). **a** and **b** are from ERA40 and **c** is from GPCP. **d-f** as **a-c** but for the CMIP2 MRI model. **g-i** as for **a-c** but for the IPCC MRI model. In **a** the boxes show the tropical and North American domains over which pattern correlations, root mean square differences, and spatial standard deviations are calculated

**Fig. 4.** Taylor diagram of December-February warm event composite surface air temperature anomalies. The plot summarizes the pattern correlation, root mean square difference, and spatial standard de-

viation of each of the CMIP2 and IPCC models and NCEP/NCAR reanalysis with respect to ERA40 over the tropics (30$^{o}$N-30$^{o}$S, 90$^{o}$E-90$^{o}$W; see the black box in **Fig. 3a**)

**Fig. 5.** Pattern correlations over the tropics (30$^{o}$N-30$^{o}$S, 90$^{o}$E-90$^{o}$W; see the black box in **Fig. 3a**) of December-February warm event composite anomlaies for the IPCC models with respect to observations are plotted against those of their CMIP2 counterparts for **a** surface air temperature (with respect to ERA40) **b** sea-level pressure (with respect to ERA40), and **c** precipitation rate (with respect to GPCP). The solid filled circles give the results based on the composites from the full integration periods of the IPCC models. The box and whisker plots give the spread of pattern correlations from the IPCC models when the data are sampled in 80 year segments that overlap by all but 10 years. The 80 year segments were chosen to match the length of the CMIP2 realizations. The extremes of the whiskers indicate the largest and smallest IPCC pattern correlations, the lower and upper limits of the boxes correspond to the 25th and 75th percentiles, and the dividing line represents the median correlation. The widths of the boxes are proportional to the number of 80 year segments. The diagonal lines represent equal pattern correlations for the CMIP2 and IPCC models

**Fig. 6.** Pattern correlations of December-February warm event composite precipitation anomalies are plotted against pattern correlations of the December-February precipitation climatology for the CMIP2 and IPCC models. The pattern correlations are calculated with respect to GPCP precipitation over the tropics (30$^{o}$N-30$^{o}$S, 90$^{o}$E-90$^{o}$W; see the black box in **Fig. 3a**). The linear regression fits for different sets of models are given along with the standard error in parentheses

**Fig. 7.** Taylor diagram of December-February warm event composite anomalies over the North American region (20$^{o}$N-70$^{o}$N, 140$^{o}$W-60$^{o}$W; see the red box in Fig. 3a) from the CMIP2 and IPCC models for **a** surface air temperature anomalies, **b** sea-level pressure anomalies, and **c** precipitation rate anom-

alies. The plots summarize the pattern correlation, root mean square difference, and spatial standard deviation of anomalies. For **a** and **b** the metrics are calculated with respect to ERA40 reanalysis, and for **c** with respect to GPCP precipitation. In all cases, statistics from other observationally based datasets are provided for comparison

**Fig. 8.** Pattern correlations of December-February warm event composite anomalies over North America ($20^{o}$N-$70^{o}$N, $140^{o}$W-$60^{o}$W) are plotted with respect to the pattern correlation of December-February warm event composite precipitation anomalies in the tropics ($30^{o}$N-$30^{o}$S, $90^{o}$E-$90^{o}$W) of **a** surface air temperature, **b** Sea-level pressure, and **c** Precipitation for the IPCC models. The lettered circles are for the full record length of the IPCC simulations while the associated colored dots are for 80 year segments that overlap by all but 10 years. For **a** and **b** the pattern correlations are calculated with respect to ERA40 and NCEP/NCAR reanalyses, and for **c** with respect to GPCP precipitation. The linear regression fit is given along with the standard error in parentheses

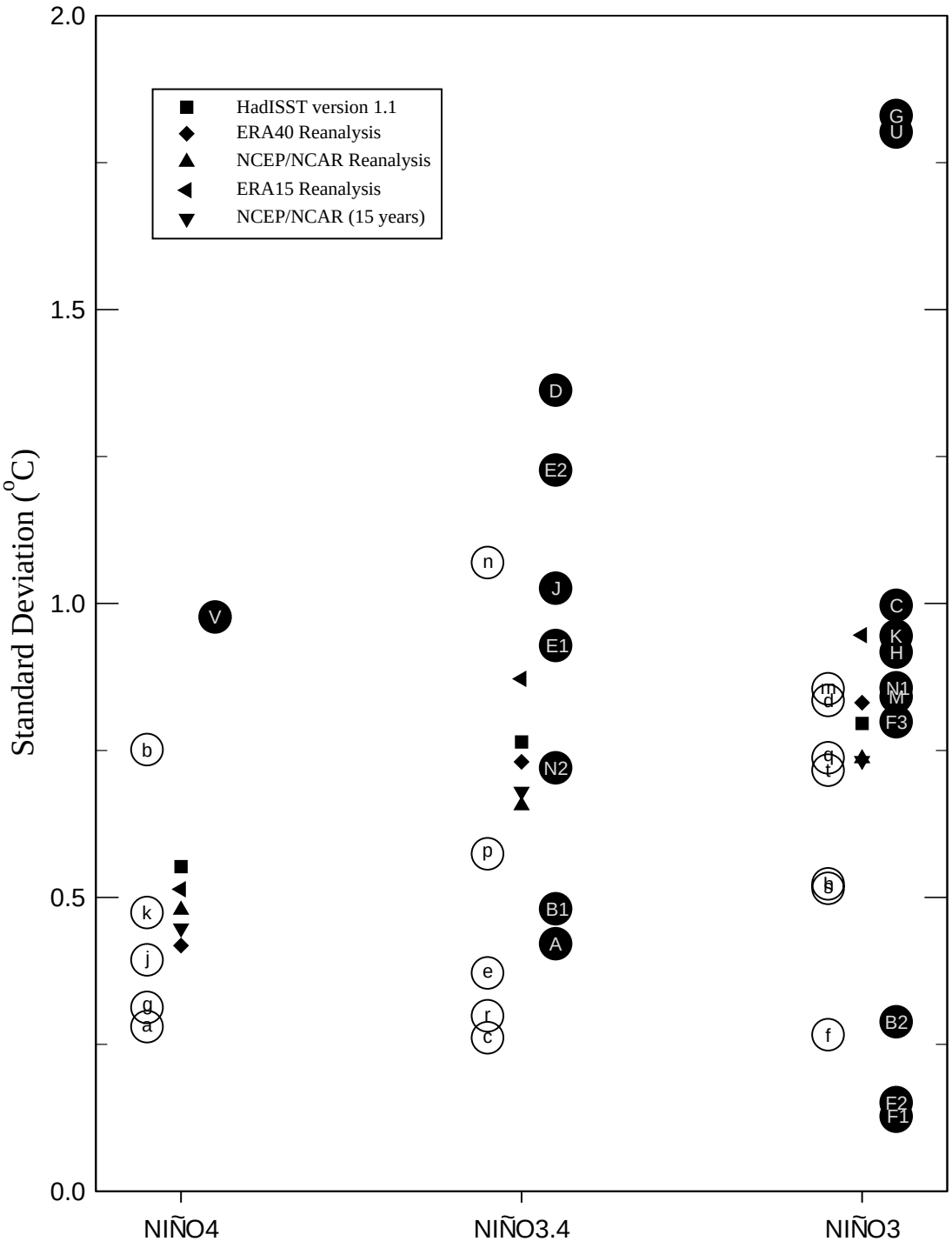


**Fig. 1.** Indices of monthly anomalies of surface air temperature ($^{o}$C) have been calculated for the NINO3 (5$^{o}$N-5$^{o}$S, 150$^{o}$W-90$^{o}$W), NINO3.4 (5$^{o}$N-5$^{o}$S, 170$^{o}$W-120$^{o}$W), and NINO4 (5$^{o}$N-5$^{o}$S, 160$^{o}$E-150$^{o}$W) regions. For the models we only plot the standard deviation for the index with the largest standard deviation. Results from the CMIP2 (IPCC) models are given by the unshaded (shaded) circles. Estimates of the observed standard deviation for each index are given from numerous sources, indicating that the observed standard deviation is largest in NINO3 and smallest in NINO4. Note: The standard deviation from HadISST is for sea surface temperature. See Table 1 for model identification notation

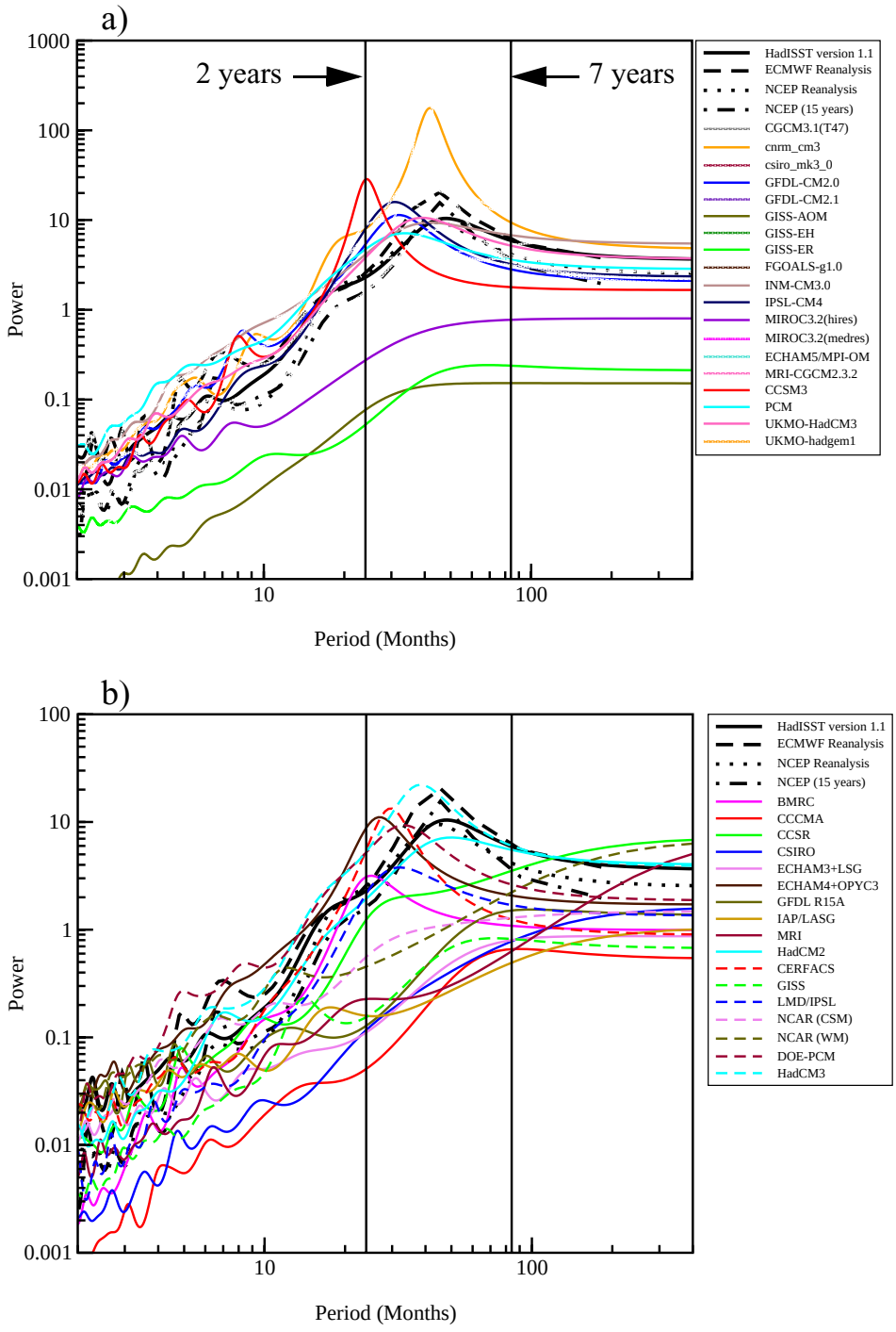


**Fig. 2.** Maximum entropy power spectra of NINO3 (5°N-5°S, 150°W-90°W) surface air temperature for **a** the IPCC models and **b** the CMIP2 models (after Fig. 4 in AchutaRao and Sperber 2002). The vertical lines correspond to periods of 2 and 7 years. The power spectra from the reanalyses and for SST from the HadISST version 1.1 dataset are given by the series of solid, dashed, and dotted black curves

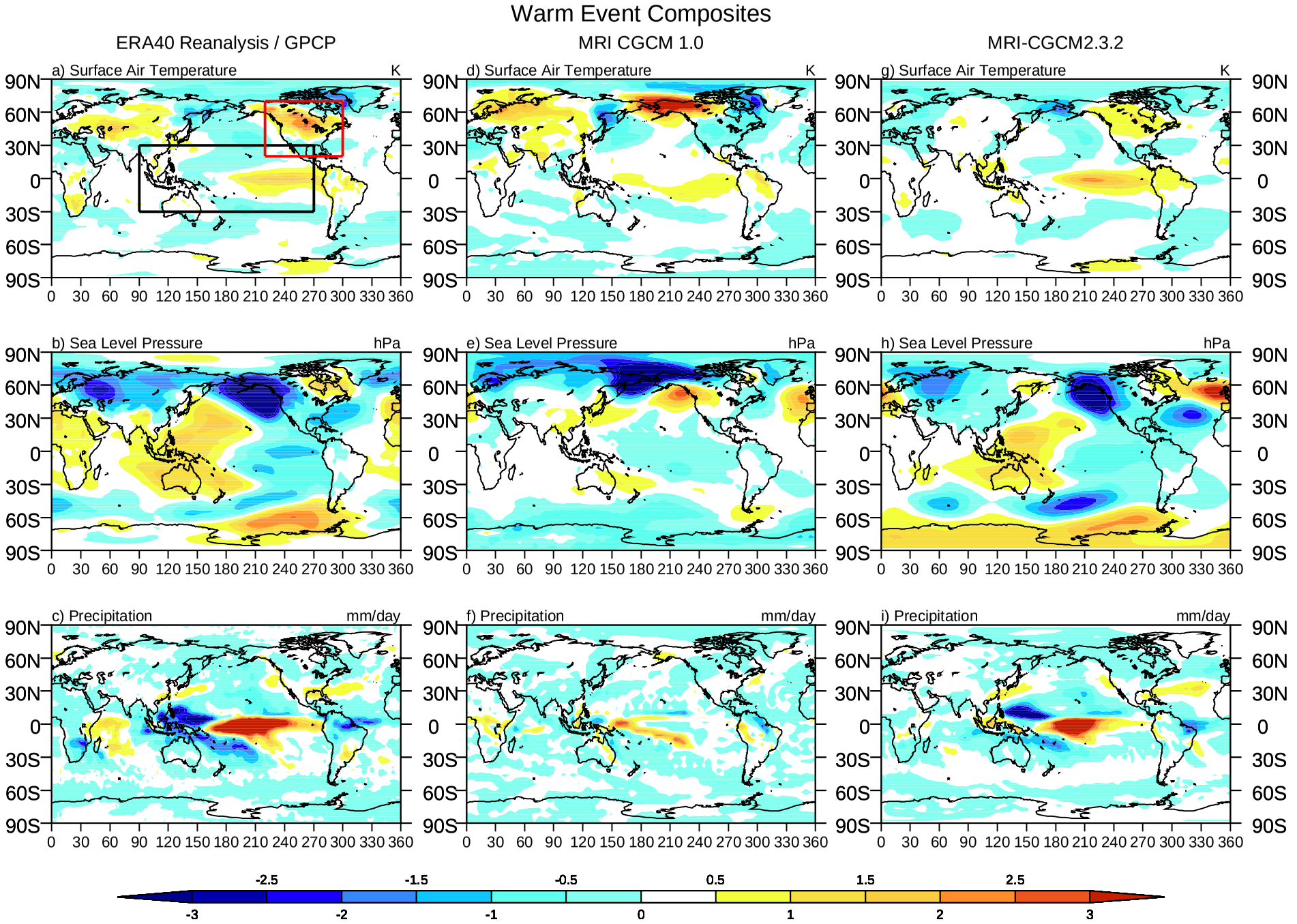


**Fig. 3.** December-February warm event composite anomalies of **a** surface air temperature ($^{o}$C), **b** sea-level pressure (hPa), **c** precipitation rate (mm day$^{-1}$). **a** and **b** are from ERA40 and **c** is from GPCP. **d-f** as **a-c** but for the CMIP2 MRI model. **g-i** as for **a-c** but for the IPCC MRI model. In **a** the boxes show the tropical and North American domains over which pattern correlations, root mean square differences, and spatial standard deviations are calculated

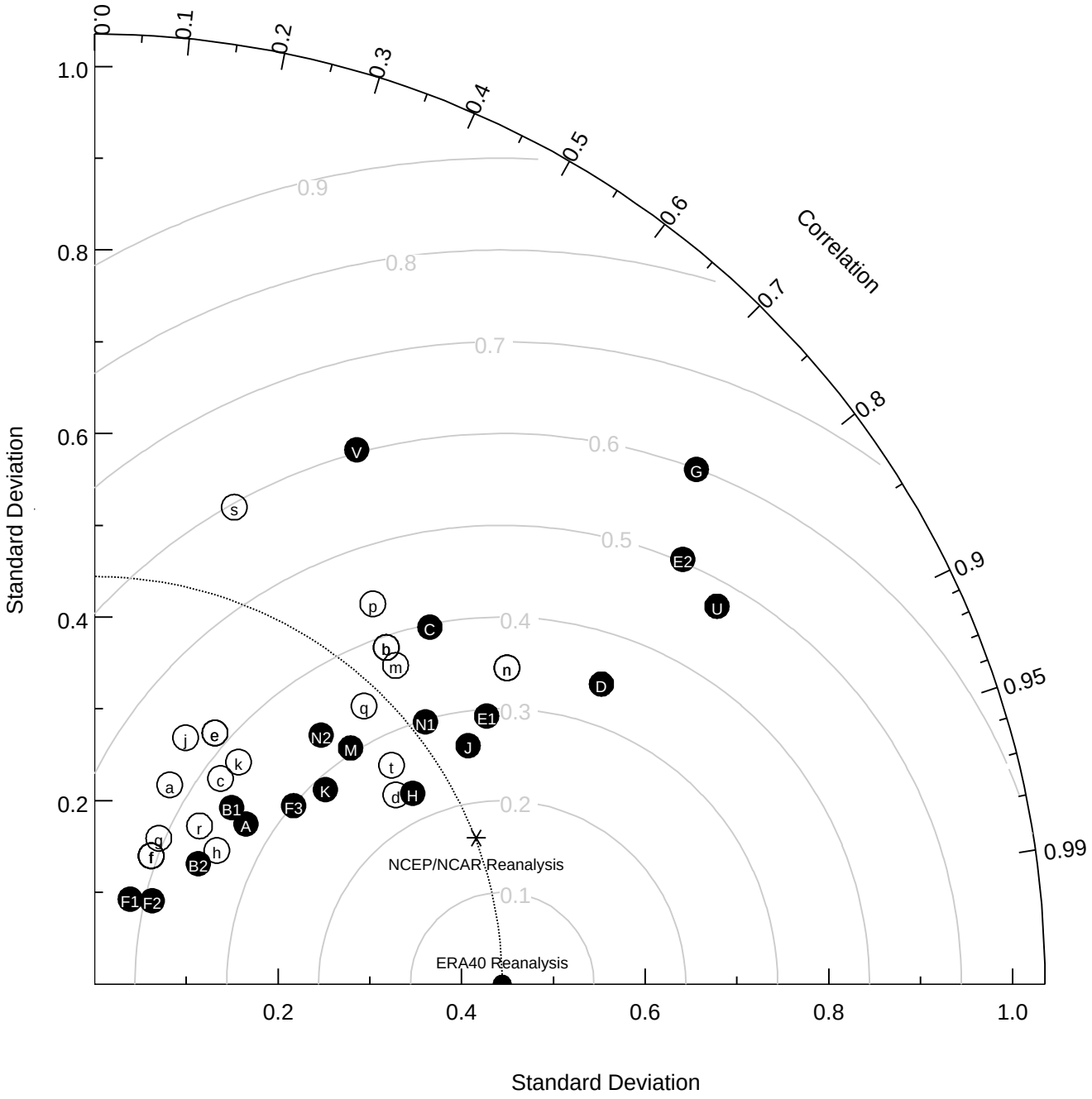


**Fig. 4.** Taylor diagram of December-February warm event composite surface air temperature anomalies. The plot summarizes the pattern correlation, root mean square difference, and spatial standard deviation of each of the CMIP2 and IPCC models and NCEP/NCAR reanalysis with respect to ERA40 over the tropics (30$^{o}$N-30$^{o}$S, 90$^{o}$E-90$^{o}$W; see the black box in **Fig. 3a**)

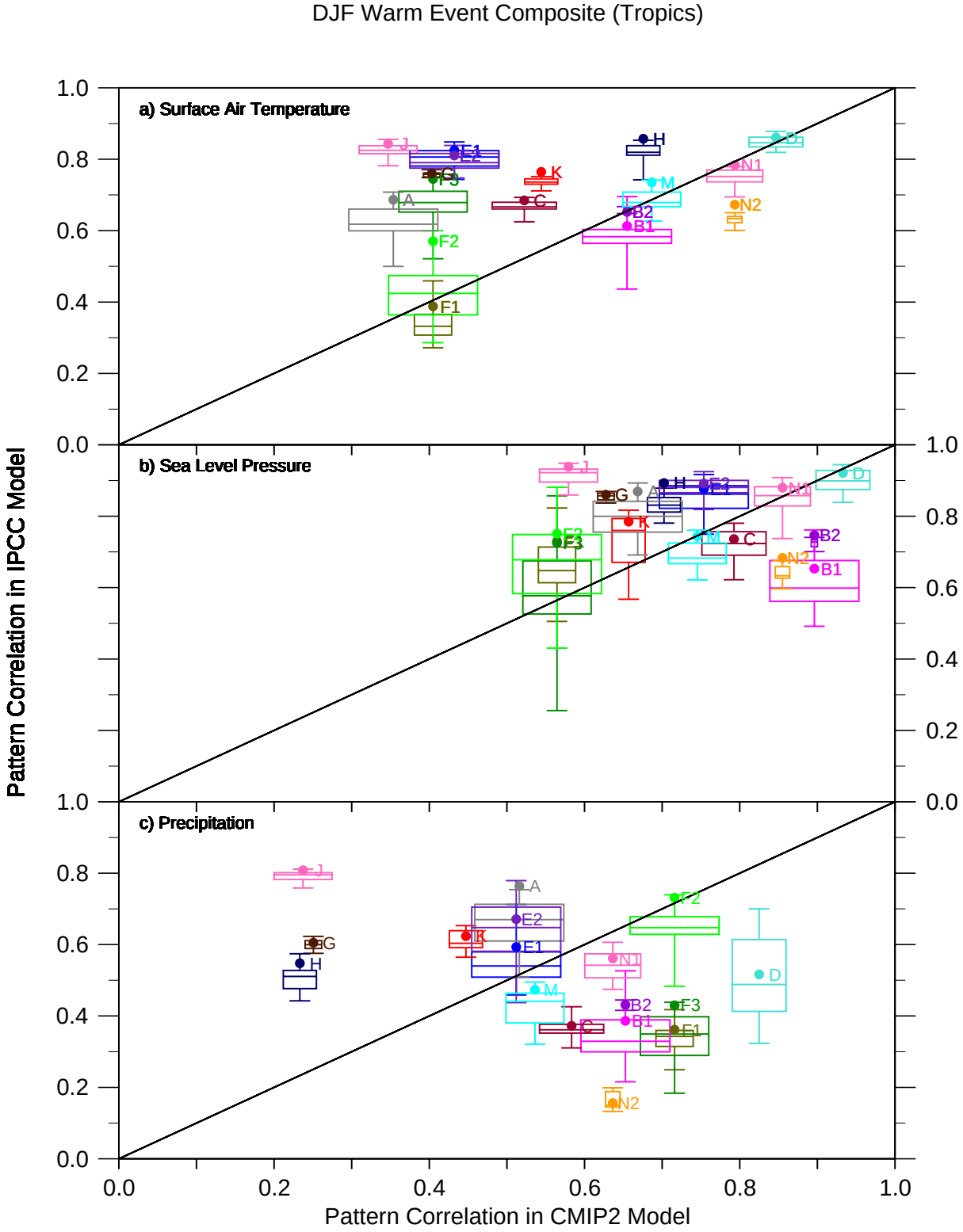


**Fig. 5.** Pattern correlations over the tropics (30°N-30°S, 90°E-90°W; see the black box in **Fig. 3a**) of December-February warm event composite anomlaies for the IPCC models with respect to observations are plotted against those of their CMIP2 counterparts for **a** surface air temperature (with respect to ERA40) **b** sea-level pressure (with respect to ERA40), and **c** precipitation rate (with respect to GPCP). The solid filled circles give the results based on the composites from the full integration periods of the IPCC models. The box and whisker plots give the spread of pattern correlations from the IPCC models when the data are sampled in 80 year segments that overlap by all but 10 years. The 80 year segments were chosen to match the length of the CMIP2 realizations. The extremes of the whiskers indicate the largest and smallest IPCC pattern correlations, the lower and upper limits of the boxes correspond to the 25th and 75th percentiles, and the dividing line represents the median correlation. The widths of the boxes are proportional to the number of 80 year segments. The diagonal lines represent equal pattern correlations for the CMIP2 and IPCC models

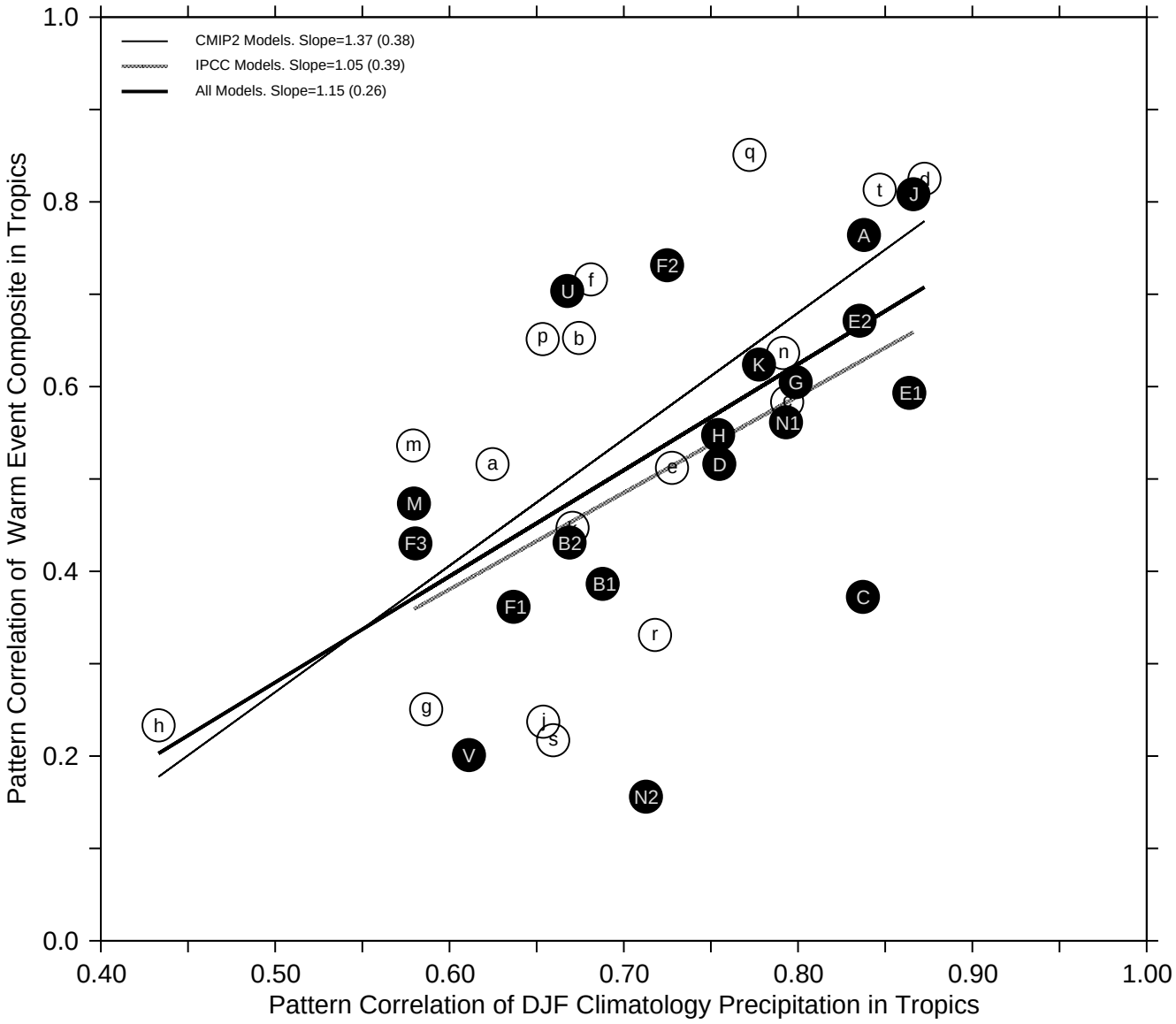


**Fig. 6.** Pattern correlations of December-February warm event composite precipitation anomalies are plotted against pattern correlations of the December-February precipitation climatology for the CMIP2 and IPCC models. The pattern correlations are calculated with respect to GPCP precipitation over the tropics ($30^{o}$N-$30^{o}$S, $90^{o}$E-$90^{o}$W; see the black box in **Fig. 3a**). The linear regression fits for different sets of models are given along with the standard error in parentheses

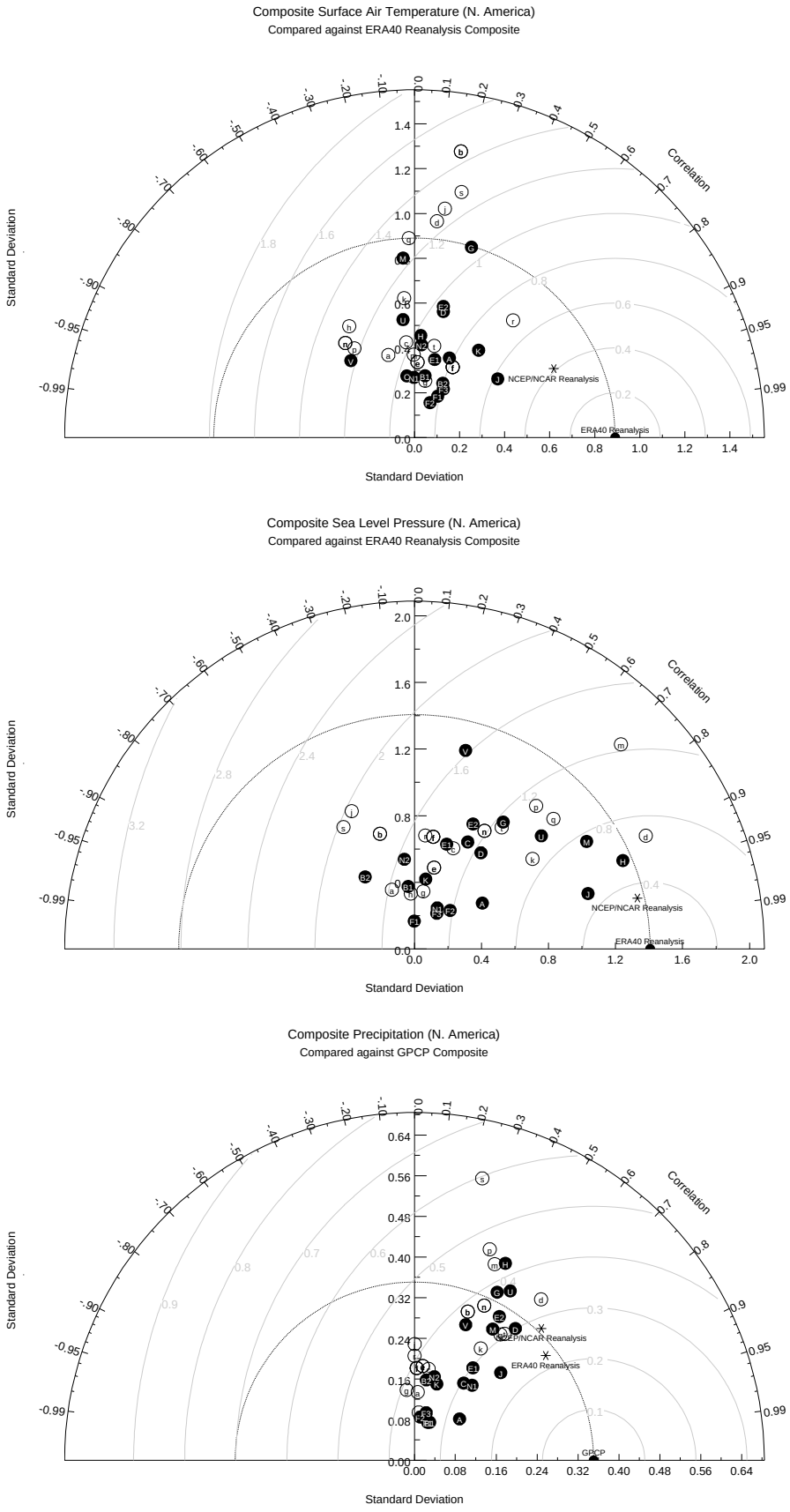


**Fig. 7.** Taylor diagram of December-February warm event composite anomalies over the North American region (20$^{o}$N-70$^{o}$N, 140$^{o}$W-60$^{o}$W; see the red box in Fig. 3a) from the CMIP2 and IPCC models for **a** surface air temperature anomalies, **b** sea-level pressure anomalies, and **c** precipitation rate anomalies. The plots summarize the pattern correlation, root mean square difference, and spatial standard deviation of anomalies. For **a** and **b** the metrics are calculated with respect to ERA40 reanalysis, and for **c** with respect to GPCP precipitation. In all cases, statistics from other observationally based datasets are provided for comparison

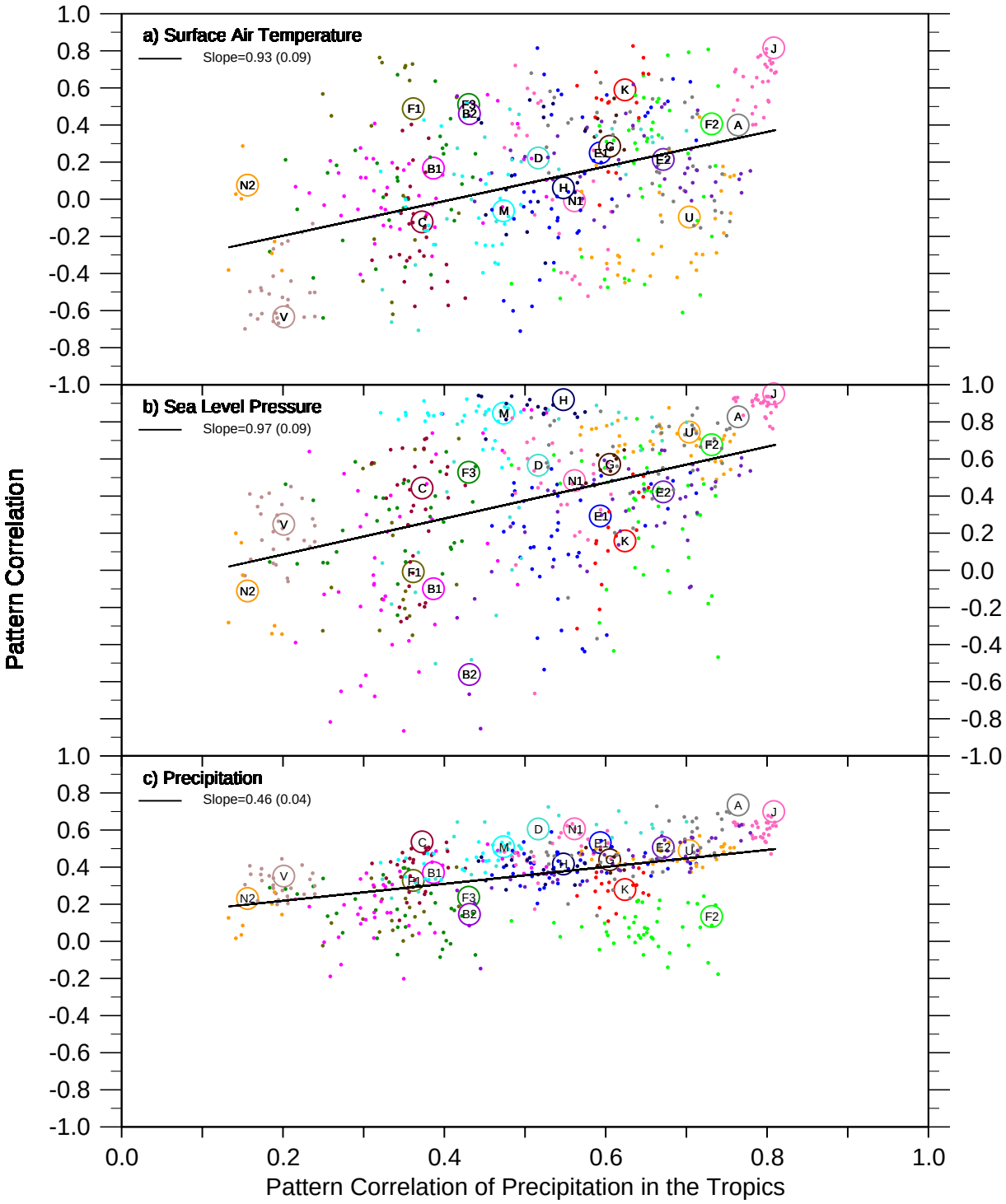


**Fig. 8.** Pattern correlations of December-February warm event composite anomalies over North America (20°N-70°N, 140°W-60°W) are plotted with respect to the pattern correlation of December-February warm event composite precipitation anomalies in the tropics (30°N-30°S, 90°E-90°W) of **a** surface air temperature, **b** Sea-level pressure, and **c** Precipitation for the IPCC models. The lettered circles are for the full record length of the IPCC simulations while the associated colored dots are for 80 year segments that overlap by all but 10 years. For **a** and **b** the pattern correlations are calculated with respect to ERA40 and NCEP/NCAR reanalyses, and for **c** with respect to GPCP precipitation. The linear regression fit is given along with the standard error in parentheses