Distinguishing the roles of natural and anthropogenically forced decadal climate variability: Implications for prediction

US CLIVAR Decadal Predictability Working Group

Amy Solomon[1,2]*, Lisa Goddard[3], Arun Kumar[4], James Carton[5], Clara Deser[7], Ichiro Fukumori[8], Arthur Greene[3], Gabriele Hegerl[9], Ben Kirtman[10], Yochanan Kushnir[11], Matthew Newman[1,2], Doug Smith[12], Dan Vimont[14], Tom Delworth[6], Jerry Meehl[7], and Timothy Stockdale[13]

[1]University of Colorado, USA
[2]Earth System Research Laboratory, NOAA, USA
[3]International Research Institute for Climate and Society, USA
[4]Climate Prediction Center, NOAA, USA
[5]University of Maryland, USA
[6]Geophysical Fluid Dynamics Laboratory, NOAA, USA
[7]National Center for Atmospheric Research, USA
[8]Jet Propulsion Laboratory, NASA, USA
[9]University of Edinburgh, UK
[10]University of Miami and Center for Ocean-Land-Atmosphere Studies, USA
[11]Lamont Doherty Earth Observatory of Columbia University, USA
[12]Met Office Hadley Centre, UK
[13]European Centre for Medium-Range Weather Forecasts, UK
[14]University of Wisconsin, USA

For Bulletin of the American Meteorological Society
Submitted November 2009

Capsule: In decadal forecasts, the magnitude of natural decadal variations may rival that of anthropogenically forced climate change on regional scales. To assess these forecasts, it is necessary to identify what processes contribute to the skill of a decadal prediction and how to distinguish between natural and externally forced variations.

*Corresponding author: Amy Solomon, NOAA/ESRL/PSD, 325 Broadway, R/PSD1, Boulder, CO 80305-3328. amy.solomon@noaa.gov

Abstract: Given that over the course of a 10-30 year forecast the magnitude of natural decadal variations may rival that of anthropogenically forced climate change on regional scales, it is envisioned that initialized decadal predictions will provide important information for climate-related management and adaptation decisions. Such predictions are presently one of the grand challenges for the climate community. Long experience in weather and climate forecasting has shown that forecasts are incomplete without a priori assessment of potential forecast skill and forecast reliability. For decadal predictions, this requires identifying those physical phenomena -- and their model equivalents or lack thereof -- that provide additional predictability and/or cause forecast spread on decadal time scales, including an assessment of the physical processes through which anthropogenic forcing interacts with or projects upon natural variability. Such a physical framework is necessary to provide a consistent assessment (and insight into potential improvement) of the decadal prediction experiments planned for the AR5.

1. Motivation

An ambitious effort to produce near-term, decadal forecasts has begun, motivated by the possibility that climate models initialized with ocean observations can capture not only the impact of the changing atmospheric composition but also the slow natural variations of the climate system. In the cases where initialization improves the forecast, addressing the question of how much that improvement is due to the natural versus the forced climate components is critical to understanding the benefits of the initialized decadal prediction effort. Untangling the natural and forced components of the climate is also necessary because the response to external forcing may project onto or comingle with natural climate variability. In particular, as the science of decadal prediction is in its infancy, one would like to assess and understand (a) the expectations for added regional climate information and skill achievable from initialized decadal predictions; (b) what physical processes or modes of variability are important to the decadal prediction and predictability problem, and whether their relevance may evolve and change with time; (c) What elements of the observing system are important for initializing and verifying decadal predictions; and (d) in terms of attribution, to what extent are regional changes in the current climate due to natural climate variations and is thus transitory, and to what extent are they due to anthropogenic forcing, and are likely to continue. Alternatively, one could question to what extent natural climate variations are masking or amplifying climate change due to anthropogenic forcing.

As with the preceding decade, the climate experienced in the near-term will be a combination of forced climate change and natural variability. As an example, consider the prolonged drought conditions of the American West since the late 1990s. Most of the

21st century climate change projections used in the AR4 suggest that this region will get drier as precipitation decreases and evaporative demand increases with future warmer temperatures (Seager et al. 2007). However, since dry conditions in this part of the world are also associated with *natural* interannual-to-decadal variability in sea surface temperatures in both the Atlantic and the Pacific basins (e.g. McCabe et al. 2004; Seager et al. 2005; Schubert et al. 2009), how much of the recent drought can be attributed to natural variability and how much can be attributed to on-going climate change? An answer to this question could greatly aid western water resource managers in developing informed adaptation strategies.

The purpose of this paper is to establish a framework within which the problem of separating decadal natural variability from anthropogenically forced variability may be systematically addressed. Note that separating decadal natural variability from anthropogenically forced variability goes beyond what has already been accomplished in previous studies that focus primarily on determining a long-term anthropogenic signal (Hegerl et al. 2007), because on decadal time scales anthropogenic effects may be non-monotonic, regionally dependent, and/or convolved with natural variability.

Towards establishing a framework to distinguish the roles of natural and anthropogenically forced decadal climate variability we address the following questions: On decadal time scales, how can we estimate the relative amplitudes, and spatial structures, of natural and forced variability? What approaches can be used to separate natural decadal variability from anthropogenically forced decadal variations? How do the results of the analysis depend on the specific method? What are the limitations for the observational data and model simulations? In addressing these questions, we must also

consider: How does the fingerprint of forced variability interact with the natural variability?

The Fifth Assessment Report of the IPCC will contain a set of decadal prediction experiments (Taylor et al. 2008). It must be clearly emphasized that these are initial experiments to assess the current feasibility of decadal predictions and will be conducted by modeling centers around the world. The detailed procedures for these prediction experiments are left up to each modeling group, for example, how to initialize the models, what observational data to use for initialization, how many ensemble members to run, and how to present the results (Meehl et al. 2009b). The approaches for separating natural and forced variability discussed in this paper, together with their benefits and limitations, are intended to serve as a starting point for the assessment of the upcoming decadal prediction experiments and to better understand the processes and potential predictability of decadal variations generally.

2. Physical processes involved with decadal time scales in the climate system

To assess naturally occurring decadal variability in the climate system and the ability of models to simulate and forecast it, one must identify the relevant physical processes. Most studies point to oceanic processes as central to climate memory, particularly those related to reservoirs of ocean heat or momentum and related atmospheric feedbacks. For example, in midlatitudes sea surface temperatures (SSTs) are well described by the stochastic climate model paradigm (Frankignoul and Hasselmann 1977), where random atmospheric surface forcing with equivalent power at all frequencies (a "white noise" spectrum) is integrated by the ocean mixed layer to produce

a spectrum which amplifies the power at lower frequencies (a "red noise" spectrum) (see Deser et al. 2009a for review).

Separating the regional natural variations from the externally forced signal is necessary to answer the next question, that is, whether the natural variability could provide additional predictability beyond external forcing or is unpredictable and merely a source of climate noise on these timescales. A number of ocean processes, for example, overturning and gyre circulations, the triggering of Rossby waves, and advection of temperature/salinity anomalies by the mean currents, may provide additional predictability by influencing atmospheric variability at the air-sea interface across large distances and long time scales.

One example of potentially predictable natural climate variability is produced by wind-forced extratropical ocean Rossby waves that propagate across an ocean basin and create thermocline anomalies near the western boundary, which are then communicated to the surface through wintertime heat fluxes and wind stress. Schneider and Miller (2001) demonstrate that such a process in the North Pacific can yield predictable wintertime SST anomalies in the Kuroshio–Oyashio Extension at lead times of up to 3 years. If these SST anomalies can create a large enough atmospheric response such that a wind stress pattern of similar structure and opposite sign is produced in the eastern part of the ocean basin, then a quasi-oscillatory cycle of 16-40 year time scales can be produced, although modeling studies provide conflicting results on this point (For example, see Schneider et al. 2002).

Another source of predictability may come from the shallow wind-driven meridional overturning ocean circulations called subtropical cells (STCs), which connect

the subtropical atmosphere to the equatorial region through the ocean in both the Atlantic and Pacific basins (see Schott et al. 2004 for review). STCs have been hypothesized to play a role in decadal climate variability by the advection of salinity/temperature anomalies along STC pathways to the equator (Gu and Philander 1997; Yeager and Large 2004) or by changes in STC strength, which controls the amount of cold water that upwells at the equator, in models (Kleeman et al. 1999; Solomon et al. 2003) and observations (McPhaden and Zhang 2002). The decadal time scale in this variability can come from the atmospheric variability at the poleward edge of the STCs, or the time it takes anomalies to travel to the equator or the spin-up time of the cell -- both of which may provide additional predictability.

A prominent source of natural climate variability is the Atlantic Meridional Overturning Circulation (AMOC). This circulation plays a key role in climate by transporting warm upper ocean water northward in the Atlantic and replacing this water with southward flowing cold deep water generally below 1000m depth. The Atlantic also shows evidence of multi-decadal climate variations generally referred to as the Atlantic Multidecadal Oscillation (AMO, Goldenberg et al. 2001), a basin-scale signature in SST and a corresponding pattern in wintertime winds and sea level pressure. The time series corresponding to these spatial patterns shows the AMO to have completed two cycles since 1900. Similar fluctuations are also found in many coupled models (e.g. Latif et al. 2006), and while different models seem to produce fluctuations for different reasons, all seem to involve a link to the AMOC. The presence of feedbacks linking AMOC, SST and the atmospheric circulation opens up potential for predictability associated with AMOC variations (see Knight et al. 2006).

It has also been recognized that anthropogenic forcing may cause the aforementioned processes to vary in space and time. For example, it has been suggested that anthropogenic forcing caused a slowing of the Pacific STCs between the 1970's-1990's (McPhaden and Zhang 2002). Also, in the Atlantic, evidence from observational studies (see Hurrell et al. 2006) suggests a link between the AMO and AMOC and anthropogenic climate change. In most coupled models the link between increasing greenhouse gasses and decreasing AMOC is evident as well (Figure 1), where a weakening of the AMOC corresponds to a reduction in salinity at higher latitudes due to greater mid-latitude precipitation in a warmer world.

3. Approaches to separate natural from anthropogenically forced decadal variations

3.1. Analysis of model ensemble means and variance

An ensemble of climate simulations can be used to identify the response in the climate system to external forcing and the variations that are internal to the system. The former is referred to as external variability, while the latter is referred to as internal or the natural climate variability. The approach described closely follows a similar approach used in seasonal climate predictions, where seasonal atmospheric variability is decomposed into external variability due to sea surface temperature (SST) and internal variability due to atmospheric processes alone (e.g. Kumar and Hoerling 1995; Rowell et al. 1995).

For coupled global circulation models (CGCM) used in climate change projections where the time evolution of external forcings (for example, CO_2, solar variability, volcanic aerosols etc.) is specified, the mean over an ensemble of CGCM

simulations is the variability that can be attributed to the specified external forcing, and the departure in each climate simulation from the ensemble mean is the internal variability due to natural fluctuations. Thus for uninitialized climate change projections, the ensemble mean represents the models' realization of the anthropogenically-forced component of the climate. This approach can be applied to any time-average extending from seasonal to annual to decadal time scales.

For large enough ensembles with specified external forcing, this approach also allows for the investigation of dominant modes of coupled variability, both internally and externally forced. Analyses using both long CGCM control simulations that do not include time-varying external forcing and of individual ensemble members relative to the ensemble mean for the time-varying external forcing also reveal the magnitude and spatial structure of natural climate variability in the model. For example, using twin experiments with and without time-varying external forcing, Meehl et al. (2009b) found that the Interdecadal Pacific Oscillation (IPO) (Power et al. 1999) emerged as a preferred mode for both internally and externally forced variability over the Pacific region.

In another example, a 40-member ensemble of CGCM integrations with changing atmospheric composition and ozone recovery for the period 2005-2060 is compared against a long (10,000 year) unforced control run of the atmospheric model component with a specified repeating annual cycle of sea surface temperatures and sea ice conditions and no changes in atmospheric composition (Deser et al. 2009b). The ensemble mean atmospheric circulation trend, which we interpret as the forced response, exhibits a statistically significant weakening of the Southern Hemisphere polar vortex during austral summer (positive sea level pressure trends at high latitudes and negative ones at middle

latitudes: Figure 2, left panel). The spread of the response among the individual ensemble members, obtained from the leading Empirical Orthogonal Function (EOF) of the 40 individual trend maps, is also characterized by an annular pattern reminiscent of the forced response (Figure 2, middle panel). Further, the pattern of the spread closely resembles the leading EOF of 56-year trends from the unforced control run (Figure 2, right panel).

The probability density functions (pdfs) of the the trends, obtained by projecting the individual trend maps from the 40-member coupled ensemble and the 10,000 yr control integration onto the leading EOF of the atmospheric control integration (Figure 2, bottom panels) show that: 1) external forcing produces a statistically significant shift in the mean of the pdf; and 2) the spread of the trends amongst the individual coupled model ensemble members can be entirely accounted for by internal atmospheric variability or "climate noise". These results demonstrate that identification of externally forced multi-decadal trends can be subject to large uncertainties owing to noise, thus requiring very large ensembles.

This approach to separating the natural and externally-forced variability, based on ensembles of climate simulations, is a conceptually simple and elegant methodology in its formulation. The approach, however, also has some limitations including the fact that estimates of internal and external variability are model dependent and may not be realistic. On the other hand, based on an analysis of simulations from multiple CGCMs, and a comparison of total variability against the observed estimates, some confidence in the model-based estimates can be gained. However, the approach, by construction, requires a large ensemble of simulations and can be computationally taxing. For

example, existing model archives as part of the CMIP3, in general, do not have large enough ensembles from individual models for this approach to be a viable.

3.2. Application of detection/attribution studies

Climate change detection and attribution studies aim to distinguish the anthropogenically-forced component of the climate. They generally use information about the shape of the expected climate response to forcing (the 'fingerprint') and are targeted to isolating the role of these fingerprints in observed near-term climate change as clearly as possible from internal climate variability. Often, this is done using signal separation techniques, such as 'optimal fingerprints' or best linear unbiased estimators (see review in Hegerl et al. 2007). For attribution, all relevant external influences on climate must be considered. The attribution methods then attempt, with uncertainty estimates, to identify the contribution of each external forcing factor to the observed change. The shape of the fingerprints is hereby assumed known, and their magnitude is estimated, allowing the results to account for uncertainties such as errors in a model's climate sensitivity to a particular forcing, or in the magnitude of external forcings in general.

The results from attribution methods, however, go further. The best guess and uncertainty ranges of the greenhouse gas contribution in the observed temperature changes can be used directly to predict future changes (Stott and Kettleborough 2002) and has been used, among other methods, to provide uncertainty ranges for future climate change in the IPCC assessment (Knutti et al. 2008). Therefore, attribution works both retrospectively (understanding past changes) and in the sense of predictions. Indeed, Lee

et al. (2006) demonstrated that over a large part of the 20th century, the forced component can produce skillful hindcasts of decadal global temperature variability.

The success of fingerprint methods in separating different factors influencing climate suggests that they may be useful to separate the influence of initial conditions from those of external forcing. While this would lead to some methodological difficulties (such as a time-evolving uncertainty ranging from atmospheric variability only to full internal climate variability at a point in time when the initial condition no longer carry any predictive capacity), such an application should be possible. This would allow for the possibility to trace where the initial conditions have made significant differences in hindcasts, and how long this influence has lasted.

However, when it comes to applying such approaches on regional scales, a number of difficulties loom. One important shortcoming is that on smaller than continental scales, the uncertainty in forcings other than greenhouse gases is large – the exact time-space pattern of aerosol forcing, land use change forcing, and other forcings is often poorly known, and poorly simulated in models. This would hamper the ability to reliably attribute successes and failures in regional hindcasts to causes. When applying to variables other than temperature, the difficulties increase dramatically. Only recently, for example, has the effect of anthropogenic forcing on precipitation been demonstrated (Zhang et al. 2007). However, the multi-model fingerprint produces smaller changes in zonal precipitation than observed, indicating that understanding and simulation of precipitation variability is still limited.

3.3. Signal to noise maximizing EOFs

The "signal to noise (S/N) maximizing EOF analysis" is a method to distinguish between the climate response to prescribed external forcing, which is common to all ensemble members, and internal (natural) climate variability, which is temporally uncorrelated between ensemble members. The response common to all ensemble members will hereafter be referred to as "the signal" and the latter as "climate noise". Note that this methodology will not help separate the anthropogenic signal from an internal signal arising from a predictable process or simple persistence that arises from the initial conditions if the latter is common to all ensemble members. This approach is important for a robust identification of a signal of unknown shape in the presence of vigorous climate noise. A variant of this approach can also enable a robust comparison between the modeled signal and the observations (particularly in the case of a hindcast) in order to determine whether the signal is present in the observations and to compare the amplitude of the signal to the unpredictable background variability.

The problem of identifying the predictable patterns can be addressed by calculating the dominant patterns (EOFs) of the covariance matrix of the ensemble-average output. The latter is taken to be a sum of two independent covariance matrices: one for the signal and the other for the climate noise. When there is spatial structure (i.e., spatial correlation) in the climate noise (an obvious assumption in this case) then the EOFs of the sum will constitute a mix between the patterns of the signal and those of the noise to a level that could be difficult for interpretation. To overcome this problem, one needs to remove the spatial structure from the noise covariance matrix (i.e., diagonalize it). To achieve this, Venzke et al. (1999) and Chang et al. (2000) project the ensemble

mean on the leading EOFs of the covariance matrix derived from the pooled deviations of the ensemble member outputs from the ensemble mean. The latter matrix thus serves as the best estimate of the covariance matrix of climate noise. When the resulting, "pre-whitened", ensemble-mean covariance matrix is subjected to an EOF analysis the loading patterns become free of the influence of climate noise. As Venzke et al. (1999) demonstrate, the EOFs of the pre-whitened covariance matrix can then be applied to the pre-whitening operator to determine the patterns that maximize the ratio of ensemble-mean variance to within-ensemble variance and thus form the optimally determined patterns of forced and predictable variability.

The S/N maximizing EOFs procedure was applied in several similar applications, where the investigators needed to identify the response to external or prescribed conditions in an ensemble of GCM integrations (e.g., Terray and Cassou 2002, Tippet and Giannini 2006). Most recently, Ting et al. (2009) looked for a way to separate natural, North Atlantic, decadal and multidecadal 20th century SST variability from the change in the Basin SST due to anthropogenic forcing (and other external causes). Normally, North Atlantic natural variability, in particular Atlantic multidecadal variability (AMV), is identified as the deviation from linearly detrended observations. To improve on this simple approach, Ting et al. (2009) used the S/N maximizing EOFs procedure to determine the externally forced signal in six CMIP3 ensembles with several different CGCMs. Starting with SST deviations from, respectively, observed and each model's climatology, they showed that the decadal variations in the observed, North Atlantic average SST time series are out of the range of the different model-based estimate of the forced signal in the Basin (see Figure 3). They defined the AMV signal as

the difference between the observed North Atlantic anomaly and the multi-model average of the forced signal. The results of this study indicate that this procedure can be useful for attribution studies associated with decadal prediction process. In a complementary analysis, Ting et al. (2009, see Figure 3 there) use a more common ANOVA procedure to show that in the 20th century, the forced climate signal amounts to less than 30% of the total variance outside of the tropics (30°S-30°N), including over land areas.

In summary, S/N maximizing EOF analysis is useful for clearly identifying the forced signal in an ensemble of forced CGCM integrations, in the presence of significant levels of climate noise. The method cannot by itself provide separate information on the patterns of externally forced variability and any predictable or persistent patterns related to the initialized conditions, as those should be part of the output common to all ensemble members. The method also depends on the ability of the model to span the range of natural variability observed in nature as well as provide a bias-free simulation of forced variability. Discrepancies between models and observations along these lines should be dealt with in other ways.

3.4. Linear inverse models

An empirical technique that fits and then tests a multivariate red noise model to the data, called linear inverse modeling (LIM) (e.g., Penland and Sardeshmukh 1995), provides an excellent approximation of the observed evolution of SST anomalies. For example, a LIM constructed from the lag-1 auto-covariance matrix of *weekly* tropical anomalies (Newman et al. 2009) more faithfully reproduces the entire observed power spectrum on seasonal-to-interannual time scales of the dominant pattern of tropical SST variability, within the 95% confidence interval, than do the corresponding spectra from

virtually all ensemble members of the "20th-century" (20c3m) CMIP3 experiments (Randall et al. 2007). In addition, a LIM constructed with annual means of Pacific basin SSTs captures the power spectra of the data, including interannual and interdecadal spectral peaks that are significant relative to univariate red noise, encompasses prominent North Pacific regime shifts of the period (Newman 2007), and shows the Pacific Decadal Oscillation (PDO; Mantua et al. 1997) is not a single physical mode but a superposition of a number of processes with different dynamical origins, only one of which may have predictability greater than 2 years (Newman et al. 2003; Schneider and Cornuelle 2005; Newman 2007).

Figure 4 shows the three least damped (and most predictable) eigenmodes of the dynamical operator from the LIM determined from annual mean Pacific basin SSTs, and their corresponding time series. The third eigenmode has characteristics of a basin-wide "decadal ENSO" pattern with PDO signature in the North Pacific (e.g., Zhang et al. 1997; Deser et al. 2004), but since its decay time is far shorter than its period it does not actually represent a regular oscillation. As a result, little long-range forecast skill is associated with this eigenmode. In fact, virtually all long-range LIM predictability comes from the two leading stationary eigenmodes. The time series of the leading eigenmode has a 100-yr trend that is very unlikely, but not impossible, to have occurred by chance relative to multivariate red noise. The second eigenmode, which has no centennial trend, has a pattern somewhat similar to the multidecadal signal found by Deser et al. (2004) (see also D'Arrigo et al. 2006), although their result may additionally include a contribution from the less persistent decadal ENSO eigenmode. The combined effects of

these two eigenmodes alone dominate the observed patterns of Pacific SST trend over both the entire century and the last fifty years.

LIMs constructed from the output of each 20c3m ensemble member suggest that the models might be underestimating potential predictability of natural variability. In most model runs, the leading eigenmode is usually about as persistent as the observed leading eigenmode in Figure 4a, and appears strongly related to the long-term trend, although some structural differences exist between the observed and modeled eigenmodes (Figure 5a). However, in virtually all of the 20th century simulations the second eigenmode is not only poorly captured but is also strikingly less persistent than in the observed LIM (Figure 5b). Most GCMs additionally produce one or two eigenmodes with a period of about 10-30 yrs but very limited predictability. However, these patterns bear only a weak resemblance to the observed decadal ENSO eigenmode, and in some cases the CGCMs split this type of decadal variability into separate tropical and North Pacific eigenmodes (not shown), likely a consequence of all the GCMs having relatively weak tropical SST decadal variability and having North Pacific SSTs apparently too independent of the Tropics (Newman 2007).

The inability of the 20th century runs to reproduce a second persistent eigenmode, let alone one whose structure is similar to Figure 5b, suggests that there remains predictable decadal variability yet to be captured by the models. Unfortunately, it is unclear how anthropogenic forcing affects this analysis. Does it introduce a new type (i.e., a new mode) of variability, does it modify some of the eigenmodes, or does it select for (or against) some of the eigenmodes? These questions cannot be easily addressed, if at all, in an analysis of the limited data record alone. We might gain some insight by

comparing LIMs of the 20th century runs with their corresponding control runs. Since neither of the two leading eigenmodes from the control LIMs is very persistent (Figure 5; also Newman 2007), it is then tempting to interpret the first observed eigenmode as related to anthropogenic forcing and the second as representing natural variability. On the other hand, such interpretation is limited given the deficiencies of the overall GCM simulation of 20th century decadal variability.

3.5. Use of paleo-climate data

Paleo-climate data derived through the analysis of "proxies" can play an important role in the delineation of natural climate variability. When comparing model internal variability with paleo-climate reconstructions it is nevertheless important to keep in mind that reconstructions reflect both variability created within the climate system, and variability imposed on the system by natural forcings, such as volcanic eruptions (Tett et al., 2007; Hegerl et al., 2007). The characterization of phenomena at decadal and longer timescales requires correspondingly long records. Proxy data are valuable in part because their length may be many times that of the instrumental record, which is necessary to distinguish between regime-like and autoregressive behavior of the climate (e.g. Overland et al. 2006). Since these records cover centuries, when anthropogenic effects on climate were small to non-existent, they can be used to provide information on the spatial and temporal signatures of natural variability. Although the hypothesis has been advanced that anthropogenic influence on climate began as early as 8 kya (Ruddiman 2003, 2005), this idea has not been universally accepted (Brook 2009), and the climate of preindustrial times, as inferred from proxy records, is generally taken to be free of significant anthropogenic influence.

Climate proxies are *surrogates:* tree-ring width, relative abundance of an isotopic species in an ice core or the ratio of trace constituents in an aragonite coral skeleton, to list a few examples. For climate studies such records must first be *calibrated,* i.e., related in a quantitative manner to appropriate instrumental data over some common period of record. A particular issue of concern for detection and attribution studies is the loss of low-frequency variance when reconstructions are projected back in time, since the consequent underestimation of natural variability would increase the likelihood of false detection of an anthropogenic "fingerprint." A calibration method that preserves low-frequency variance is discussed by Hegerl et al. (2006), in a study that uses tree-ring data to estimate natural variability of northern hemisphere temperature.

Proxies tend to be associated by type with particular environments – corals with low-latitude marine settings, tree rings with temperate terrestrial sites, and so on. However, different proxy types may be combined in order to obtain more complete spatial sampling (Mann 2002). Also, inferences may be made about teleconnections or contemporaneous variability, and how those have varied in past centuries, by comparing paleo-records between regions. For example, coral-based reconstructions of tropical variability have been compared with tree-ring-derived records of the subtropics (Cole et al. 2002; D'Arrigo et al. 2005), examining such relationships as La Nina events and US drought.

Although most paleo-records consist of time series associated with single geographical points, tree-ring data in particular are fairly extensive, and can be used for climate model validation in the spatially distributed sense (e.g., Collins et al. 2002, Bradley 1999). Paleo-records can also be used to indicate the temporal character of

natural variability; Prairie et al. (2008) utilized a tree-ring based reconstruction of Colorado river streamflow to infer wet- and dry-state switching characteristics. A number of studies utilizing tree-ring data have focused on the regime-like behavior often attributed to Pacific decadal variability (PDV -- Minobe 1997; Biondi et al. 2001) or the quasi-periodic behavior of the Atlantic Multidecadal variability (AMV -- Gray et al. 2004). The records used in these studies had lengths of order 400 yr. Thus paleo-records, by providing long histories, yield information on regional characteristics of low-frequency natural variability, which may be useful in its own right or as a check or constraint in validating the natural variability simulated in models.

3.6. Analysis of initialized decadal prediction studies

Natural and forced variability may also be separated to a certain extent by comparing parallel sets of initialized and uninitialized hindcast experiments made with the same climate model (Smith et al. 2007; Keenlyside et al. 2008; Pohlmann et al. 2009). If all external forcing (i.e, from anthropogenic greenhouse gases and aerosols, solar irradiance and volcanic eruptions) is identical, then differences between the two sets of hindcasts arise purely from initialization. Since natural internal variability can only be predicted by starting from its correct phase, improved skill in initialized over uninitialized hindcasts may indicate skillful prediction of some aspects of natural variability. However, improved skill in initialized hindcasts may also arise from removing biases that exist in uninitialized climate models forced by observed changes in external forcing. This source of additional skill is potentially important for improving predictions of climate change commitment or short-term response to volcanic eruptions,

but would need to be taken into account in any attempt to separate natural and forced variability.

There are also other issues to be considered when analyzing hindcasts. Climate models cannot be initialized perfectly with incomplete observations. This usually leads to an initialization shock, during which the model rapidly adjusts to imbalances introduced by imperfect initialization, causing a degradation of forecast skill that could mask any signals from natural variability. Furthermore, initializing and assessing decadal hindcasts is severely hampered by the sparsity of historical sub-surface ocean observations. For example, natural variations of the Atlantic meridional overturning circulation (AMOC) are predictable in idealized model experiments (Collins et al. 2006), but our ability to achieve such predictability in reality, and hence identify the signal of natural variability, is compromised by the lack of historical ocean observations.

Model errors are also an important source of uncertainty in decadal forecasts (Hawkins and Sutton 2009), potentially masking any signals of natural variability. For these reasons, a lack of improved skill in initialized over uninitialized hindcasts does not necessarily imply a lack of potentially predictable natural variability. It is also possible that unrealistic model responses to imperfect observations (Acero-Schertzer et al. 1997; Ji et al. 2000; Masina et al. 2001) could lead to apparent hindcast skill that could be incorrectly attributed to natural variability. Ultimately these issues must be overcome in order to capitalize on the predictability of natural variability to improve decadal forecasts. If this can be achieved then improvements in skill arising from initialization should be consistent with the signals of natural variability identified by the other methods discussed in this section.

4. Challenges

4.1. Interaction between natural and externally-forced variability

As discussed earlier, the response to external forcing may resemble the natural modes of variability on regional and hemispheric scales. This was seen to be the case in the modeling study of Meehl et al. (2009a), where natural and externally forced patterns of variability with similar structure contributed to the mid 1970's climate shift over the Pacific basin, from relatively cool to relative warm conditions along the equator. Indeed, they argued that an anthropogenically-forced shift would have occurred in the 1960's if it were not for the presence of large amplitude natural variations that delayed the shift into the 1970's.

Just how external forcing interacts with natural modes of variability remains an important but unresolved issue. The process may be fundamentally linear with external forcing selecting certain natural internal modes due to their inherent time scales and spatial structures or nonlinear where the impact of the external forcing on the modes of variability has a net effect on the long-term trend (for example, see Branstator and Selten 2009). In the linear case, the forcing and the response may not have similar patterns due to the non-normal growth of natural modes. In the nonlinear case, the external forcing may cause changes in the frequency of occurrence of climate modes with or without changing the spatial structure of the leading modes of variability (see Corti et al 1999; Hsu and Zwiers 2001; Brandefelt 2006; Branstator and Selten 2009).

4.2. Observational uncertainties

Verification of the forced component of 20th century climate trends simulated in model experiments may be hampered by the limited sampling in both space and time of the observations and proxy records. In particular, knowledge of the spatial patterns and

magnitudes of climate trends over the oceans is hampered by the uneven and changing distribution of commercial shipping routes (Figure 6). The paucity of data over the Tropical Pacific Ocean before ~1960 is especially noteworthy given the influence of this region on global climate. Limited observational sampling in the marine record also affects the quality of globally-complete reconstructed data sets based on optimal interpolation techniques. An example of the impact of observational uncertainties on 20th century sea surface temperature (SST) trends is shown in Figure 7 based on an un-interpolated data set (HadSST2: Rayner et al., 2006) and two optimally-interpolated reconstructions (HadISST: Rayner et al., 2003; and ERSSTv3: Smith et al., 2008). Although trends from the three data sets share many features in common, such as a strengthening of the equatorial Pacific zonal temperature gradient (Karnauskas et al. 2009), there are also differences most notably the eastern equatorial Pacific which shows cooling in HadISST and warming in HadSST2 and ERSSTv3 (see also Vecchi et al., 2008). In addition, non-physical ship track signatures are discernable in the trends based on HadSST2. These observational sampling issues underscore the challenge of providing a robust target for model validation of 20th century surface marine climate trends.

4.3. Modeling uncertainties

Another challenge is that the spatial structure and dominant time scales of natural variations differs across models (see discussion of Figure 5). An additional challenge is that coupled climate models produce a range of responses, in space and time, to anthropogenic radiative forcing, as estimated by the first EOF of low-pass annual mean sea surface temperature anomalies (SSTs) from three coupled climate model simulations of the 20th Century (Figure 8). Identifying to what extent the temporal variations in the

climate models' forced signal (as seen in the principal component time series of the first EOF, Figure 8b,d,f,h) are due to the interaction between the forced signal and natural variability will improve our understanding of why decadal predictions differ across models. This understanding can be used to reduce model biases and potentially improve the skill of initialized forecasts.

In the tropical Pacific, the historical changes and future response of the mean state to imposed anthropogenic forcing has been a subject of debate. Different mechanisms disagree on the expected sign of change in the zonal SST gradient in tropical Pacific (Knutson and Manabe 1995; Cane et al. 1997; Clement and Seager 1999) in response to anthropogenic forcing. The observational record does little to clarify the situation, as trends in different observed SST records differ in even their sign (see Figure 7). Models that simulate the largest El Niño-like response also have the least realistic simulations of ENSO variability (Collins 2005). Indeed, Collins (2005) shows that when models are weighted by the fidelity with which they simulate ENSO variability, there is little change in the zonal SST gradient in the Pacific.

Different climate model responses to radiative forcing cause differences in the slowly varying base state of the oceans by changing the strength of gyres and overturning circulations and the spatial pattern of upper ocean heat content. Changes in the ocean base state in turn may alter predictability of natural variability by changing the advective time scale of density/salinity anomalies and pathways between the extratropics and tropics. Models need to reproduce the observed spatial patterns of forced variability in order to simulate basin and regional scale changes, otherwise ocean initialization will have limited usefulness. For example, anthropogenic sources are hypothesized to force

the winds that drive the North Pacific Gyre Oscillation and the concomitant nutrient concentrations that control the marine ecosystem (DiLorenzo et al. 2008). Initializing the ocean state should improve predictability due to thermal inertia in the upper ocean in the short term but without the correct winds the model will drift away from this state.

4.4. Uncertainties in ocean state initialization

The available analyses of ocean observations span a wide range of products aimed at climate studies as well as ocean nowcasting and short-term forecasting applications. The products differ in the underlying models and estimation methods that are used, as well as the suite of observations that are assimilated. Many of the analysis products span multi-decades from the 1980s to the present with some also reaching back to the 1950s, providing a convenient means for retrospective studies of climate variability. Assessing the relative accuracy and fidelity of the analyses depend in part on the particular metrics that are used and is an active area of study. While many of the syntheses employ estimation methods based on those first developed in weather forecasting, some employ so-called smoothing methods that estimate the source of the model inaccuracies corrected by the combination with data. In addition, the assimilation of Argo data in these analyses may remove biases in the upper ocean and allow for the initialization of ocean circulations and transports, for example see Forget et al. (2007).

5. Developing a framework to assess decadal predictions

Given that over the course of a 10-30 year forecast the magnitude of natural decadal variations may rival that of anthropogenically forced climate change on regional scales, it is envisioned that initialized decadal predictions will provide important information for climate-related management and adaptation decisions. Such predictions

are presently one of the grand challenges for the climate community. Long experience in weather and climate forecasting has shown that forecasts are incomplete without a priori assessment of potential (ensemble mean) forecast skill and forecast reliability. This will be no less true for decadal forecasts if they are to be useful. That is, there is very limited utility in merely *generating* a long-term forecast from some specified initial conditions and estimate of external forcing. Even crudely estimating predictability for a forecast system requires some understanding of the sources for potential skill, especially when expected skill depends upon the initial conditions themselves, and the expected forecast spread, even when such spread is unrealizable due to small ensemble size. For decadal predictions, this requires identifying those physical phenomena -- and their model equivalents or lack thereof -- that provide additional predictability and/or cause forecast spread on decadal time scales, including an assessment of the physical processes through which anthropogenic forcing interacts with or projects upon natural variability. Such a physical framework is necessary to provide a consistent assessment of the differing decadal prediction experiments planned for the AR5.

The main conclusion we draw from the body of work reviewed in this paper is that distinguishing between natural and externally forced variations is a very difficult problem that is nevertheless key to any assessment of decadal predictability. Note that all these techniques are generally limited by some assumption intrinsic to their analysis, such as the spatial characteristics of the anthropogenic signal, independence of noise from signal, or statistical stationarity; also, all the techniques utilize either short and potentially inaccurate observational datasets on the one hand and/or lengthier CGCM datasets based on flawed models on the other. There is a clear need for new techniques and more

sophisticated theoretical models developed specifically with the anthropogenic/natural variability separation problem in mind. This assumes, of course, that the anthropogenic signal *can* be separated, which at least some theories of the climate response to external forcing might suggest to be impossible. On the other hand, some hope in this regard may be found in the fairly linear decomposition of global mean temperature evolution generated by different external forcings in the AR4 20th century simulations (Hegerl et al. 2007), although it remains unclear how well this holds on regional scales. Both old and new techniques need to be simultaneously applied to both the long control CGCM runs and both the 20th and 21st century simulations, which will serve as a critical testbed for analysis of the relationship between natural and anthropogenic variability. These strategies can also be applied to existing decadal prediction experiments and climate change projections in order to develop a series of metrics that can be used to assess the predictions that will be done for the AR5. These metrics need to quantify, to the extent possible with limited ensemble sizes, the impact of different initialization strategies, model biases, and errors in model physics on the response to external forcing and the predictable and unpredictable natural variations.

A reasonable starting point for these metrics is to focus on decadal predictability due to ocean processes, as discussed in Section 2. This requires analysis that assesses the spatial patterns and associated time scales of natural variations, and their potential change in structure and frequency due to external forcing – work on this can begin by comparing the existing climate change projections against their companion control runs. In addition, since externally forced SSTs play an important role in climate variations over land through atmospheric teleconnections it is necessary to develop metrics that assess the

spatial pattern of externally forced SST variability, as well as, upper ocean structure and variability. Further, signal to noise ratios increase for integrated ocean fields, such as upper ocean heat content. To quantify signal to noise ratios it is necessary to develop metrics such as the pdfs described in Section 3.1. The development of these metrics will help guide the assessment of decadal forecasts and will provide a framework for identifying potential directions to improve our ability to make decadal predictions.

References

Acero-Schertzer, C.E., D.V. Hansen, and M.S. Swenson, 1997: Evaluation and diagnosis of surface currents in the National Centers for Environmental Prediction's ocean analyses. *J. Geophys. Res.*, **102**, 21,037–21,048.

Biondi, F., A. Gershunov, and D.R. Cayan. 2001: North Pacific decadal climate variability since AD 1661. *J. Climate,* **14**, 5-10.

Bradley, R., 1999: Paleoclimatology, 2nd Ed., Int. Geophysics Series 68, 610 pp, Academic.

Brandefelt, J., 2006: Atmospheric modes of variability in a changing climate. *J. Climate,* **19**, 5933-5942.

Branstator, G. and F.M. Selten, 2009: Modes of variability and climate change. *J. Climate*, **22**, 2639-2658.

Brook, E.J., 2009: Paleoclimate: Atmospheric carbon footprints? *Nature Geo*., **2**,170-172.

Cane, M.A., A.C. Clement, A. Kaplan, Y. Kushnir, D. Pozdnyakov, R. Seager, S.E. Zebiak and R. Murtugudde, 1997: Twentieth-century sea surface temperature trends. *Science*, **275**, 957-960.

Chang, P., R. Saravanan, L. Ji, and G.C. Hegerl, 2000: The effect of local sea surface temperatures on atmospheric circulation over the tropical Atlantic sector. *J. Climate*, **13,** 2195-2216.

Clement, A.C., and R. Seager, 1999: Climate and the tropical oceans. *J. Climate*, **12**, 3383-3401.

Cole, J.E., and J.T. Overpeck, 2002: Multiyear La Niña events and persistent drought in the contiguous United States. *Geophys. Res. Lett*., **29**, doi:10.1029/2001GL013561.

Collins, M., 2005: El Nino- Or La Nina-Like Climate Change? *Clim. Dyn*., **24**, 89-104.

Collins, M., T.J. Osborn, S.F.B. Tett, K. R. Briffa, and F.H. Schweingruber, 2002: A comparison of the variability of a climate model with paleotemperature estimates from a network of tree-ring densities. *J. Climate*, **15,** 1497-1515.

Collins, M., and coauthors, 2006: Interannual to decadal climate predictability in the North Atlantic: A multi-model ensembles study. *J. Climate*, **19**, 1195–1203.

Cook, E.R., D.M. Meko, D.W. Stahle, and M.K. Cleaveland, 1999: Drought reconstructions for the continental United States. *J. Climate*, **12**, 1145-1162.

Corti, S., F. Molteni, and T.N. Palmer, 1999: Signature of recent climate change in frequencies of natural atmospheric circulation regimes. *Nature*, **398**, 799-802.

D'Arrigo, R., R. Wilson, C. Deser, G. Wiles, E. Cook, R. Villalba, A. Tudhope, J. Cole,

and B. Linsley, 2005: Tropical-North Pacific climate linkages over the past four centuries. *J. Climate*, **18**, 5253–5265.

Deser, C., A.S. Phillips, and J.W. Hurrell, 2004: Pacific interdecadal climate variability: Linkages between the Tropics and the North Pacific during boreal winter since 1900. *J. Climate*, **17**, 3109–3124.

Deser, C., M.A. Alexander, S.-P. Xie, and A.S. Phillips, 2009a: Sea surface temperature variability: Patterns and mechanisms. *Ann. Rev. Mar. Sci*., in press.

Deser, C., V. Bourdette, G. Branstator, H. Teng, and A.S. Phillips, 2009b: Uncertainties in future climate trends: The role of weather noise. In preparation.

Deser, C., and A.S. Phillips, 2009: A re-assessment of 20th Century sea surface temperature trends over the tropical Pacific. In preparation.

DiLorenzo, E., and Coauthors, 2008: North Pacific Gyre Oscillation links ocean climate and ecosystem change. *Geophys. Res. Lett*., **35**, L08607, doi:10.1029/2007GL032838.

Feldstein, S.B., 2000: The timescale, power spectra, and climate noise properties of teleconnection patterns. *J. Climate*, **13**, 4430–4440.

Forget, G., B. Ferron, and H. Mercier, 2007: Combining ARGO profiles with a general circulation model in the North Atlantic. Part I: estimation of hydrographic and circulation anomalies from synthetic profiles, over a year. *Ocean Modell.* **20**,1–16.

Frankignoul C., and K. Hasselmann, 1977: Stochastic climate models. Part 2. Application to sea-surface temperature variability and thermocline variability. *Tellus*. **29**, 284-305.

Goldenberg, S.B., C.W. Landsea, A.M. Mestas-Nuñez, W.M. Gray, 2001: The recent increase in Atlantic hurricane activity: Causes and implications, *Science,* **293**, 474-479.

Gray, S. T., L. J. Graumlich, J. L. Betancourt, and G. T. Pederson, 2004: A tree-ring based reconstruction of the Atlantic Multidecadal Oscillation since 1567 A.D.. *Geophys. Res. Lett.,* **31**, L12205, doi:10.1029/2004GL019932.

Gu, D., and S.G.H. Philander, 1997: Interdecadal climate fluctuations that depend on exchange between the tropics and extratropics. *Science*, **275**, 805–807.

Hawkins, E., and R. Sutton, 2009: Decadal predictability of the Atlantic Ocean in a coupled GCM: Forecast skill and optimal perturbations using linear inverse modeling. *J. Climate*, **22**, 3960–3978.

Hegerl, G.C., T.J. Crowley, M. Allen, W.T. Hyde, H.N. Pollack, J. Smerdon, and E. Zorita, 2006: Detection of human influence on a new, validated 1500-year temperature reconstruction. *J. Climate*, **20**, 650-666.

Hegerl, G.C., and Coauthors, 2007: Understanding and attributing climate change. In: *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* [Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M. Tignor and

H.L. Miller (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Hurrell, J.W., and Coauthors, 2006: Atlantic climate variability and predictability: A CLIVAR perspective, *J. Climate,* **19**, 5100-5120.

Hsu, C.J., and F. Zwiers, 2001: Climate change in recurrent regimes and modes of Northern Hemisphere atmospheric variability. *J. Geophy.Res.-Atmos*, **106**, 20145-20159.

Ji, M., R.W. Reynolds, and D.W. Behringer, 2000: Use of TOPEX/Poseidon Sea Level Data for Ocean Analyses and ENSO Prediction: Some Early Results. *J. Climate*, **13**, 216–231.

Karnauskas, K.B., R. Seager, A. Kaplan, Y. Kushnir, and M.A. Cane, 2009: Observed strengthening of the zonal sea surface temperature gradient across the equatorial Pacific Ocean, *Journal of Climate*, **22**, 4316-4321.

Keenlyside, N., M. Latif, J. Jungclaus, L. Kornblueh, and E. Roeckner, 2008: Advancing decadal-scale climate prediction in the North Atlantic Sector, *Nature*, **453**, 84-88.

Kleeman, R., J.P. McCreary, and B.A. Klinger, 1999: A mechanism for generating ENSO decadal variability. *Geophys. Res. Lett.*, **26**, 1743–1746.

Knight, J.R., C. K. Folland, A.A. Scaife, 2006: Climate impacts of the Atlantic Multidecadal Oscillation. *Geophys. Res. Lett.*, **33**, L17706, doi:10.1029/2006GL026242.

Knutson, T.R., and S. Manabe, 1995: Time-mean response over the tropical Pacific to increased CO_2 in a coupled ocean–atmosphere model. *J. Climate*, **8**, 2181–2199.

Knutti, R., and Coauthors, 2008: A review of uncertainties in global temperature projections over the twenty-first century. *J. Climate*, **21**, 2651-2663.

Kumar, A. and M.P. Hoerling, 1995: Prospects and limitations of seasonal atmospheric GCM predictions. *Bull. Ame. Meteor. Soc.,* **76,** 335-345.

Latif, M., M Collins, H. Pohlmann, and N. Keenlyside, 2006: Predictability studies of Atlantic Sector climate on decadal time scales. *J. Climate*, **19**, 5971–5987.

Lee, T.C.K., F.W. Zwiers, X. Zhang, and M. Tsao, 2006: Evidence of decadal climate prediction skill resulting from changes in anthropogenic forcing. *J. Climate,* **19**, 5305-5318.

Mann, M.E., 2002: The value of multiple proxies. *Science*, **5586**, 1481-1482.

McCabe, G., M. Palecki, and J. Betancourt, 2004: Pacific and Atlantic influences on multidecadal drought frequency in the US. *PNAS*, **101**, 4136-4141.

Madden, R.A., 1976: Estimates of the natural variability of time-averaged sea-level pressure. *Mon. Wea. Rev*., **104**, 942–952.

Mantua, N.J., S.R. Hare, Y. Zhang, J.M. Wallace, and R.C. Francis, 1997: A Pacific interdecadal climate oscillation with impacts on salmon production. *Bull. Amer. Meteor.*

Soc., **78**, 1069–1079.

Masina, S., N. Pinardi, and A. Navarra, 2001: A global ocean temperature and altimeter data assimilation system for studies of climate variability. *Clim. Dyn.*, **17**, 687-700.

McPhaden, M.J., and D. Zhang, 2002: Showdown of the meridional overturning circulation in the upper Pacific Ocean. *Nature*, **415**, 603-608.

Meehl, G.A., and Coauthors, 2007: Global Climate Projections. In: *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* [Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M. Tignor and H.L. Miller (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Meehl, G.A., A. Hu, and B.D. Santer, 2009a: The mid-1970s climate shift in the Pacific and the relative roles of forced versus inherent decadal variability *J. Climate,* **22**, 780-792.

Meehl, G.A., and Coauthors, 2009b: Decadal prediction: Can it be skillful?, *Bull. Amer. Meteor. Soc.*, DOI: 10.1175/2009BAMS2778.1 (available through AMS early on-line releases).

Minobe, S., 1997: A 50–70 year climatic oscillation over the North Pacific and North America. *Geophys. Res. Lett.*, **24**, 683–686.

Newman, M., 2007: Interannual to decadal predictability of tropical and North Pacific sea

surface temperatures. *J. Climate*, **20**, 2333–2356.

Newman, M., G.P. Compo, and M.A. Alexander, 2003: ENSO-forced variability of the Pacific decadal oscillation. *J. Climate*, **16**, 3853–3857.

Newman, M., P.D. Sardeshmukh, and C. Penland, 2009: How important Is air–sea coupling in ENSO and MJO evolution? *J. Climate*, **22**, 2958–2977.

Overland, J.E., D.B. Percival, and H.O. Mofjeld, 2006: Regime shifts and red noise in the North Pacific. *Deep-Sea Res*., **53**, 582-588.

Penland, C., and P.D. Sardeshmukh, 1995: The optimal growth of tropical sea surface temperature anomalies. *J. Climate*, **8**, 1999–2024.

Pohlmann, H., J. Jungclaus, J. Marotzke, A. Köhl, and D. Stammer, 2009: Improving Predictability through the Initialization of a Coupled Climate Model with Global Oceanic Reanalysis. *J. Climate*, **22**, 3926–3938.

Power, S., T. Casey, C. Folland, A. Colman, and V. Mehta, 1999: Interdecadal modulation of the impact of ENSO on Australia. *Clim. Dyn*., **15**, 319--324.

Prairie, J., K. Nowak, B. Rajagopalan, U. Lall, and T. Fulp, 2008: A stochastic nonparametric approach for streamflow generation combining observational and paleoreconstructed data. *Water Resour. Res*., **44**, doi:10.1029/2007WR006684.

Randall, D.A., and Coauthors, 2007: Cilmate Models and Their Evaluation. In: *Climate*

Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change [Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M.Tignor and H.L. Miller (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Rayner, N.A., D.E. Parker, E.B. Horton, C.K. Folland, L.V. Alexander, D.P. Rowell, E.C. Kent, and A. Kaplan, 2003: Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *J. Geophys. Res*., **108**, 4407.

Rayner, N.A., P. Brohan, D.E. Parker, C.F. Folland, J.J. Kennedy, M. Vanicek, T. Ansell, and S.F.B. Tett, 2006: Improved analyses of changes and uncertainties in sea surface temperature measured in situ since the mid-nineteenth century: the HadSST2 data set. *J. Climate*, **19**, 446-469.

Rowell, D.P., C.K. Folland, K. Maskell, and M.N. Ward, 1995: Variability of summer rainfall over tropical North Africa (1906-92): Observations and modelling, *Q. J. R. Meteorol. Soc*., **121**, 669-704.

Ruddiman, W.F., 2003: The anthropogenic greenhouse era began thousands of years ago. *Clim. Change*, **61,** 261-293.

Ruddiman, W.F., 2005: The early anthropogenic hypothesis a year later. *Clim. Change*, **69**, 427-434.

Schneider, N., and A.J. Miller, 2001: Predicting western North Pacific Ocean climate. *J. Climate*, **14**, 3997–4002.

Schneider, N., and B.D. Cornuelle, 2005: The forcing of the Pacific Decadal Oscillation. *J. Climate*, **18**, 4355–4373.

Schneider, N., A.J. Miller, and D.W. Pierce, 2002: Anatomy of North Pacific decadal variability. *J. Climate*, **15**, 586–605.

Schott, F. A., J.P. McCreary, and G.C. Johnson, 2004: Shallow overturning circulations of the tropical subtropical oceans. *Earth Climate: The Ocean–Atmosphere Interaction, Geophys. Monogr.,* **147**, Amer. Geophys. Union, 261–304.

Seager, R., and Coauthors, 2007: Model projections of an imminent transition to a more arid climate in southwestern North America. *Science*, **316**, 1181-1184.

Seager, R., Y. Kushnir, C. Herweijer, N. Naik, and J. Velez, 2005: Modeling of tropical forcing of persistent droughts and pluvials over western North America: 1856–2000. *J. Climate*, **18**, 4065–4088.

Schubert, S., and Coauthors, 2009: A USCLIVAR project to assess and compare the responses of global climate models to drought-related SST forcing patterns: Overview and results, *J. Climate*, DOI: 10.1175/2009JCLI3060.1 (available through AMS early on-line releases).

Smith, D., S. Cusack, A. Colman, A. Folland, G Harris, and J. Murphy, 2007: Improved surface temperature prediction for the coming decade from a global circulation model. *Science*, **317**, 796-799.

Smith, T.M., R.W. Reynolds, T.C. Peterson, and J. Lawrimore, 2008: Improvements to NOAA's historical merged land-ocean surface temperature analysis (1880-2006). *J. Climate*, **21**, 2283-2296.

Solomon, A., J.P. McCreary, R. Kleeman, and B.A. Klinger, 2003: Interactions between interannual tropical oscillations and decadal extratropical oscillations in an intermediate coupled model of the Pacific basin. *J. Climate*, **16**, 2395-2410.

Stott P.A., and J.A. Kettleborough, 2002: Origins and estimates of uncertainty in twenty-first century temperature rise. *Nature*, **416**, 723–726.

Stott, P.A., G.S. Jones, J.A. Lowe, P. Thorne, C. Durman, T.C. Johns, and J.C. Thelen, 2006: Transient climate simulations with the HadGEM1 climate model: Causes of past warming and future climate change. *J. Climate*, **19**, 2763–2782.

Taylor, K.E., R.J. Stouffer, and G.A. Meehl, 2008: A summary of the CMIP5 Experimental Design. http://www-pcmdi.llnl.gov/.

Tett, S.F.B., and Coauthors, 2007: The impact of natural and anthropogenic forcings on climate and hydrology since 1550. *Clim. Dyn*., **28**, 3-34.

Terray, L., and C. Cassou, 2002: Tropical Atlantic sea surface temperature forcing of quasi-decadal climate variability over the North Atlantic-European region. *J. Climate*, **15,** 3170-3187.

Ting, M.F., Y. Kushnir, R. Seager, and C.H. Li, 2009: Forced and internal Twentieth-Century SST trends in the North Atlantic. *J. Climate*, **22,** 1469-1481.

Tippett, M.K., and A. Giannini, 2006: Potentially predictable components of African summer rainfall in an SST-forced GCM simulation. *J. Climate*, **19**, 3133-3144.

Vecchi, G.A., A.C. Clement, and B.J. Soden, 2008: Examining the tropical Pacific's response to global warming. *EOS*, **89**, 81-83.

Venzke, S., M.R. Allen, R.T. Sutton, and D.P. Rowell, 1999: The atmospheric response over the North Atlantic to decadal changes in sea surface temperature. *J. Climate*, **12**, 2562-2584.

Wunsch, C., 1999: The interpretation of short climate records, with comments on the North Atlantic and Southern Oscillations. *Bull. Amer. Meteor. Soc*, **80**, 245–255.

Yeager, S.G., and W.G. Large, 2004: Late-winter generation of spiciness on subducted isopycnals. *J. Phys. Oceanogr.*, **34**, 1528–1547.

Zhang, X., F.W. Zwiers, G.C. Hegerl, N. Gillett, H. Lambert, S. Solomon, P. Stott and T. Nozawa, 2007: Detection of human influence on 20th Century precipitation trends. *Nature*, **468**, 461-466.

Zhang, Y., J.M. Wallace, and D.S. Battisti, 1997: ENSO-like interdecadal variability. *J. Climate*, **10,** 1004–1020.

FIGURE CAPTIONS

Figure 1. Strength of the AMOC at 30^{o}N in a variety of 19 AR4 coupled models forced with observed greenhouse gas and aerosol forcing until 1999 and the SRES A1B scenario of greenhouse gas forcing after 1999. Bars on the left show various observational estimates. From Meehl et al. (2007).

Figure 2. Projected November-February sea level pressure trends during 2005-2060 over the Southern Hemisphere. *Top left*: Forced 40-member coupled model ensemble mean. *Top middle*: Leading EOF of the deviation of each coupled model ensemble member's trend from the coupled model ensemble mean trend. *Top right*: Leading EOF of a 178-member ensemble of 56-yr trends from a 10,000 yr atmospheric model control integration. *Bottom left*: PDF of the trends in the index of the Southern Annular Mode from each coupled model ensemble member (red bars) and from each atmospheric control member (gray). *Bottom right*: As in bottom left, but the coupled model ensemble mean trend has been removed from each individual coupled model ensemble member. From Deser et al. (2009b).

Figure 3. (a) Projection of SST averaged in the North Atlantic Basin onto the leading S/N-maximizing PC in each of the participating models (see list in Figure and information in Ting et al., 2009). Each model PC is depicted by a different color and the dashed line is the ensemble averaged. The observed SST average, suggesting a superposition of a forced trend and internal, multi-decadal variability, is shown in the

solid black line. (b) The observed AMO index constructed by subtracting from the observed North Atlantic SST average the model estimates of the forced NA SST shown in the top panel. The black dashed line in the bottom panel is the determined from the forced response average across all six participating models. From Ting et al. (2009).

Figure 4. Leading empirical eigenmodes and their corresponding time series (right) from the LIM of annual-mean HadISST SST anomalies. The LIM is constructed as in Newman (2007) except the EOF basis is determined over the entire Pacific domain (20^{o}S-60^{o}N); the leading 12 PCs are retained, explaining 92% of the variance in both the tropics and in the North Pacific, unlike Newman where under two-thirds of the North Pacific variance was retained. Contour interval is the same in all panels but is arbitrary. Red (blue) shading indicates positive (negative) values; zero contour is removed for clarity. a) Leading eigenmode, stationary with decay time of 13 yrs. b) Second eigenmode, stationary with decay time of 6.4 yrs. c) Most energetic phase of third ("decadal ENSO") eigenmode, propagating with period 16 yrs and decay time of 2.1 yrs.

Figure 5. Comparison between the a) leading and b) second observed eigenmodes with the corresponding eigenmodes based on each 100-yr ensemble member from the 20^{th} century AR4 coupled GCMs (blue) and the associated control runs (green). Both plots show the decay time scale of each modeled eigenmode vs. its pattern correlation with the corresponding observed eigenmode. The red circle in each panel indicates the observed eigenmode.

Figure 6. Distribution of surface marine observations from the International Comprehensive Ocean-Atmosphere Data Set, shown as the percent of months with at least 1 observation per 2º latitude x 2º longitude grid box during the 20 year period indicated. Adapted from Deser et al. (2009a).

Figure 7. Twentieth century sea surface temperature trends (C 100 yr^{-1}) from the un-interpolated HadSST2 (top), reconstructed HadISST (middle), and reconstructed ERSSTv3 datasets, based on monthly anomalies during 1900-2008. A minimum of 3 months/decade in each decade was required to compute a trend from the HadSST2 data set. From Deser and Phillips (2009).

Figure 8. First empirical orthogonal function (EOF) and associated principal component of annual mean sea surface temperature from observations and three 20^{th} Century simulations for years 1890-1999. (A,B) The Hadley Centre sea surface temperature data set (HadISST, Rayner et al., 2003). (C,D) The National Center for Atmospheric Research/Department of Energy Parallel Climate Model Version 1 (NCAR/PCM1, http://www.cgd.ucar.edu/pcm/). (E,F) The Geophysical Fluid Dynamics Laboratory Climate Model version 2.1 (GFDL/CM2.1, Delworth et al. 2006). (G,H) The National Center for Atmospheric Research Community Climate System Model version 3.0 (NCAR/CCSM3.0, http://www.ccsm.ucar.edu/models/ccsm3.0). All data has been smooth with a 10-year low-pass Lanczos filter using 21 weights. EOF patterns are normalized. Principal components are in units of °C. The percent in the upper right of each figure indicates the amount of variance explained by each pattern. Note that the

principal component time series from the climate model simulations show fluctuations with larger amplitude than observations, all of which fluctuate on different time scales.

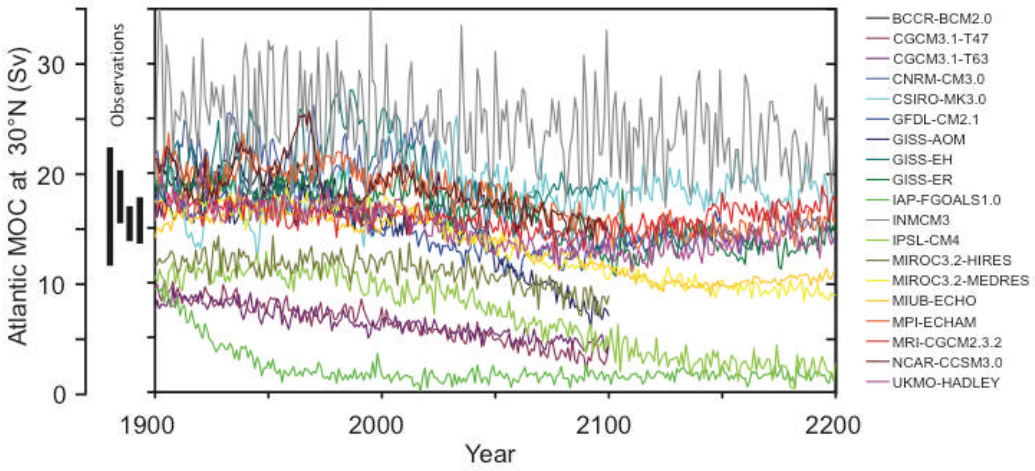
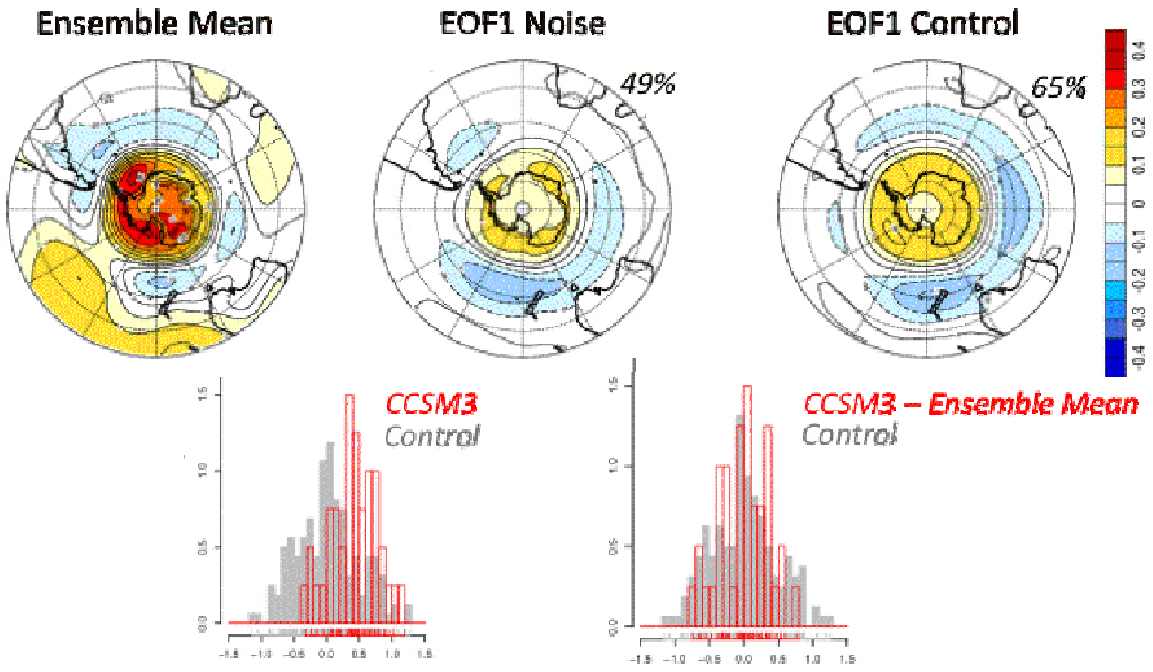


Figure 1. Strength of the AMOC at 30°N in a variety of 19 AR4 coupled models forced with observed greenhouse gas and aerosol forcing until 1999 and the SRES A1B scenario of greenhouse gas forcing after 1999. Bars on the left show various observational estimates. From Meehl et al. (2007).



Projected November-February sea level pressure trends during 2005-2060 over the Southern Hemisphere. *Top left*: Forced 40-member coupled model ensemble mean. *Top middle*: Leading EOF of the deviation of each coupled model ensemble member's trend from the coupled model ensemble mean trend. *Top right*: Leading EOF of a 178-member ensemble of 56-yr trends from a10,000 yr atmospheric model control integration. *Bottom left*: PDF of the trends in the index of the Southern Annular Mode from each coupled model ensemble member (red bars) and from each atmospheric control member (gray). *Bottom right*: As in bottom left, but the coupled model ensemble mean trend has been removed from each individual coupled model ensemble member. From Deser et al. (2009b).

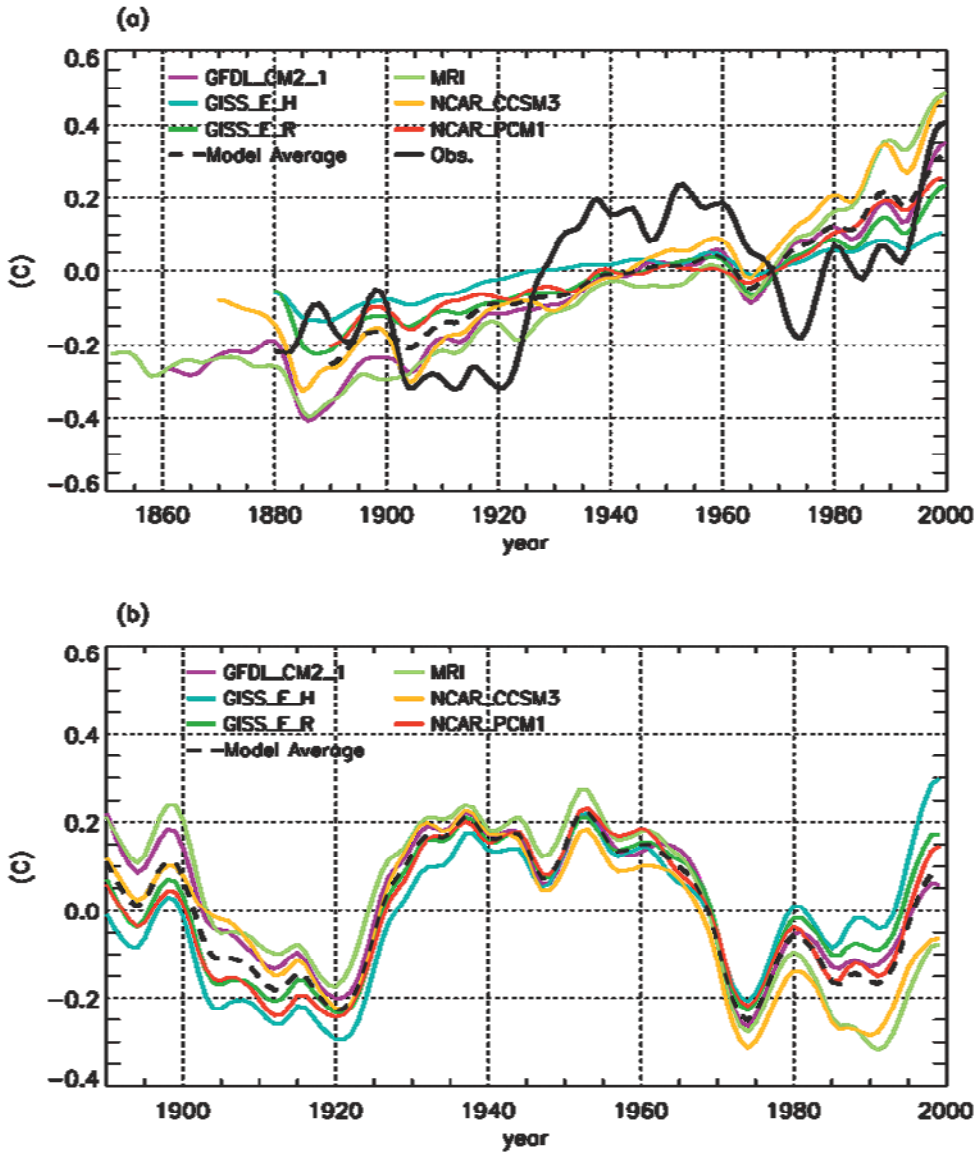


Figure 3. (a) Projection of SST averaged in the North Atlantic Basin onto the leading S/N-maximizing PC in each of the participating models (see list in Figure and information in Ting et al., 2009). Each model PC is depicted by a different color and the dashed line is the ensemble averaged. The observed SST average, suggesting a superposition of a forced trend and internal, multi-decadal variability, is shown in the solid black line. (b) The observed AMO index constructed by subtracting from the observed North Atlantic SST average the model estimates of the forced NA SST shown in the top panel. The black dashed line in the bottom panel is the determined from the forced response average across all six participating models. From Ting et al. (2009).

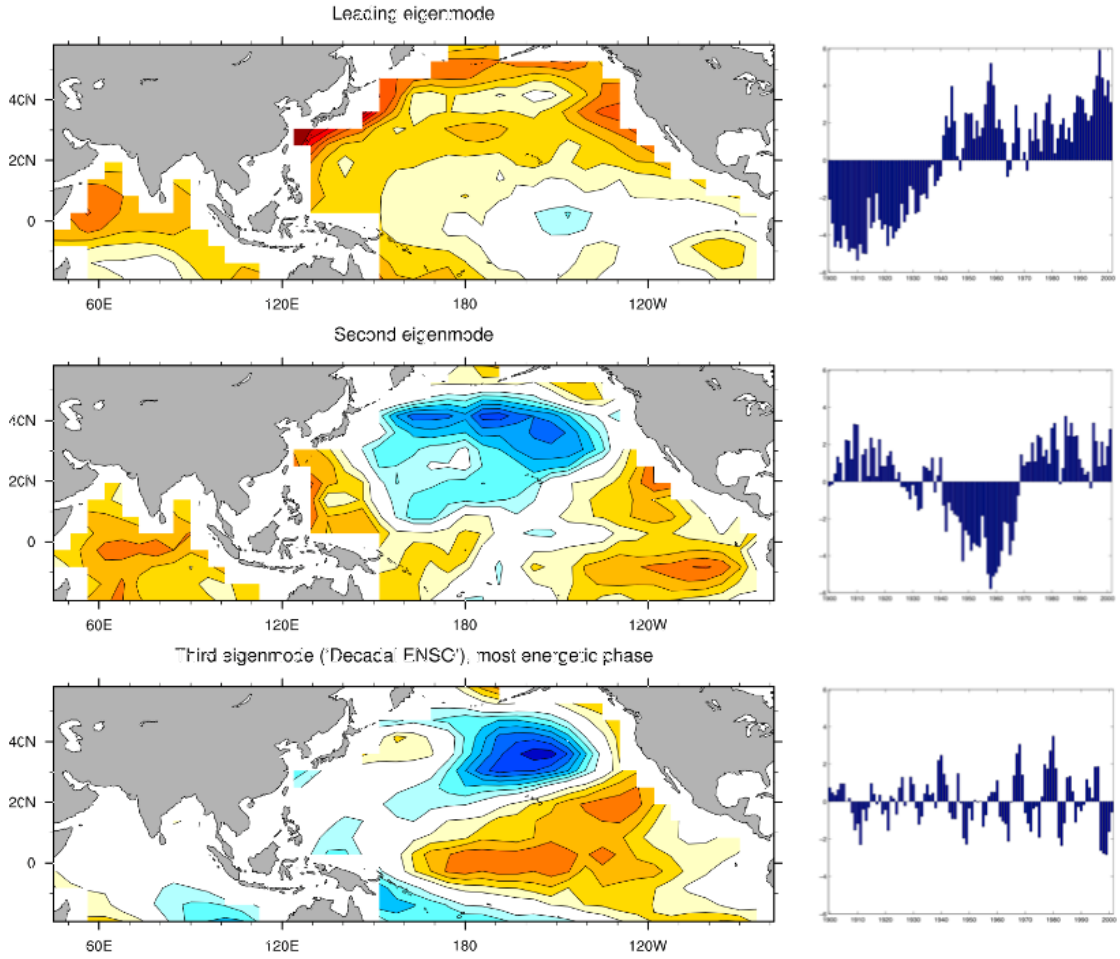


Figure 4. Leading empirical eigenmodes and their corresponding time series (right) from the LIM of annual-mean HadISST SST anomalies. The LIM is constructed as in Newman (2007) except the EOF basis is determined over the entire Pacific domain (20°S-60°N); the leading 12 PCs are retained, explaining 92% of the variance in both the tropics and in the North Pacific, unlike Newman where under two-thirds of the North Pacific variance was retained. Contour interval is the same in all panels but is arbitrary. Red (blue) shading indicates positive (negative) values; zero contour is removed for clarity. a) Leading eigenmode, stationary with decay time of 13 yrs. b) Second eigenmode, stationary with decay time of 6.4 yrs. c) Most energetic phase of third ("decadal ENSO") eigenmode, propagating with period 16 yrs and decay time of 2.1 yrs.

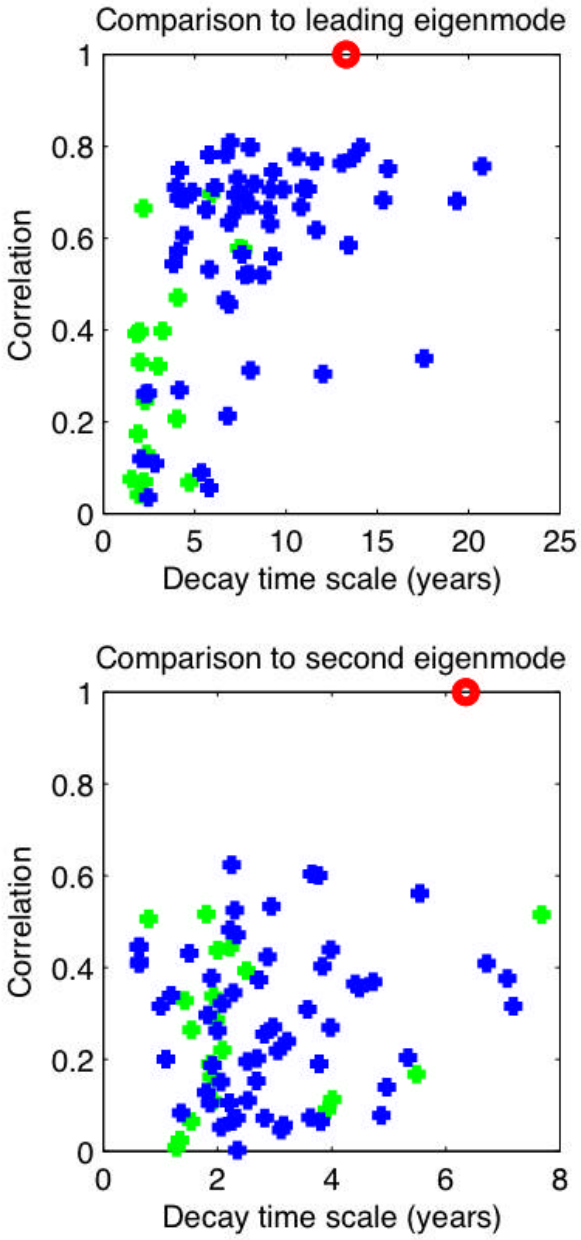


Figure 5. Comparison between the a) leading and b) second observed eigenmodes with the corresponding eigenmodes based on each 100-yr ensemble member from the 20th century AR4 coupled GCMs (blue) and the associated control runs (green). Both plots show the decay time scale of each modeled eigenmode vs. its pattern correlation with the corresponding observed eigenmode. The red circle in each panel indicates the observed eigenmode.

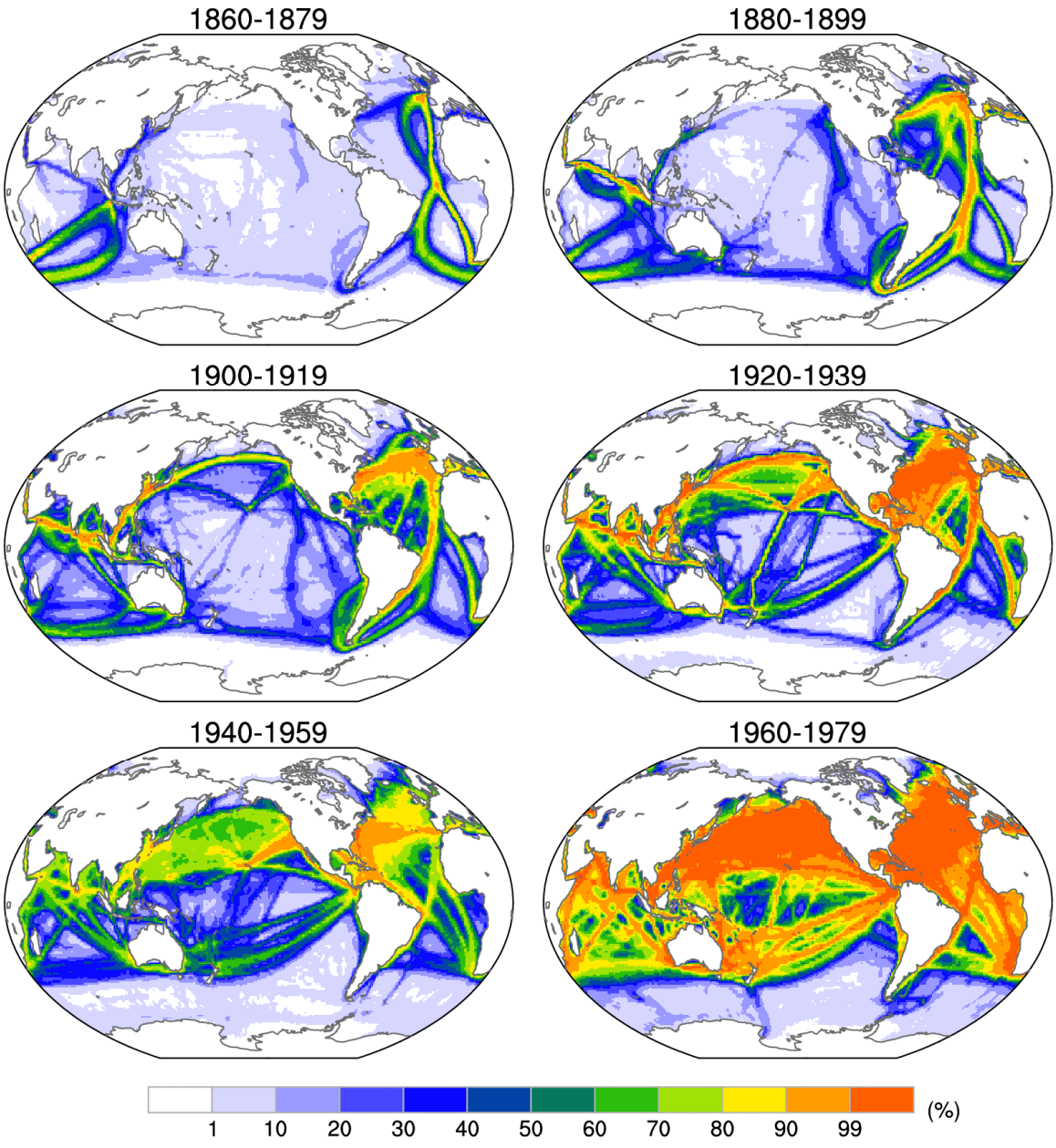


Figure 6. Distribution of surface marine observations from the International Comprehensive Ocean-Atmosphere Data Set, shown as the percent of months with at least 1 observation per 2º latitude x 2º longitude grid box during the 20 year period indicated. Adapted from Deser et al. (2009a).

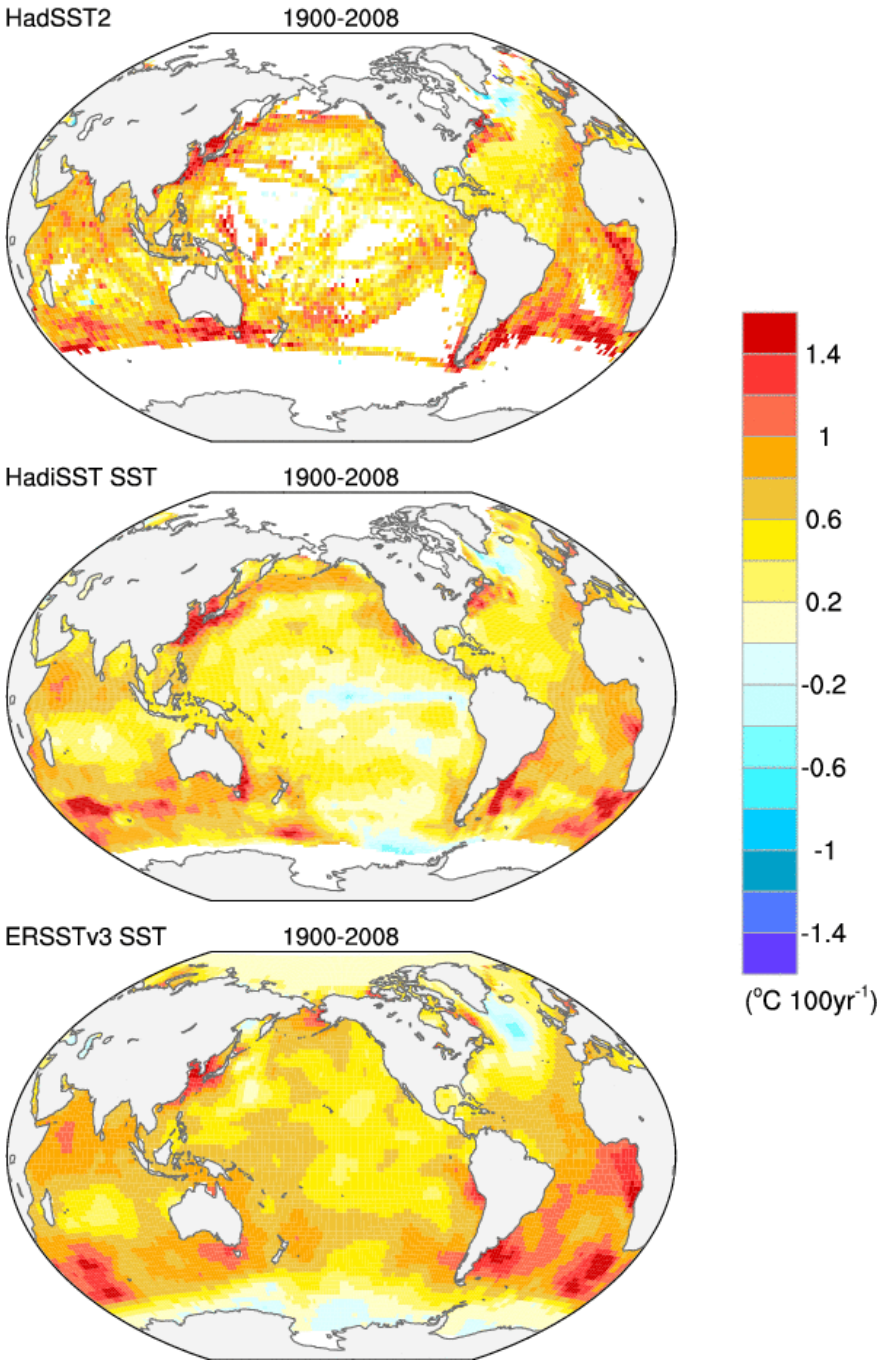


Figure 7. Twentieth century sea surface temperature trends (C 100 yr^{-1}) from the un-interpolated HadSST2 (top), reconstructed HadISST (middle), and reconstructed ERSSTv3 datasets, based on monthly anomalies during 1900-2008. A minimum of 3 months/decade in each decade was required to compute a trend from the HadSST2 data set. From Deser and Phillips (2009).

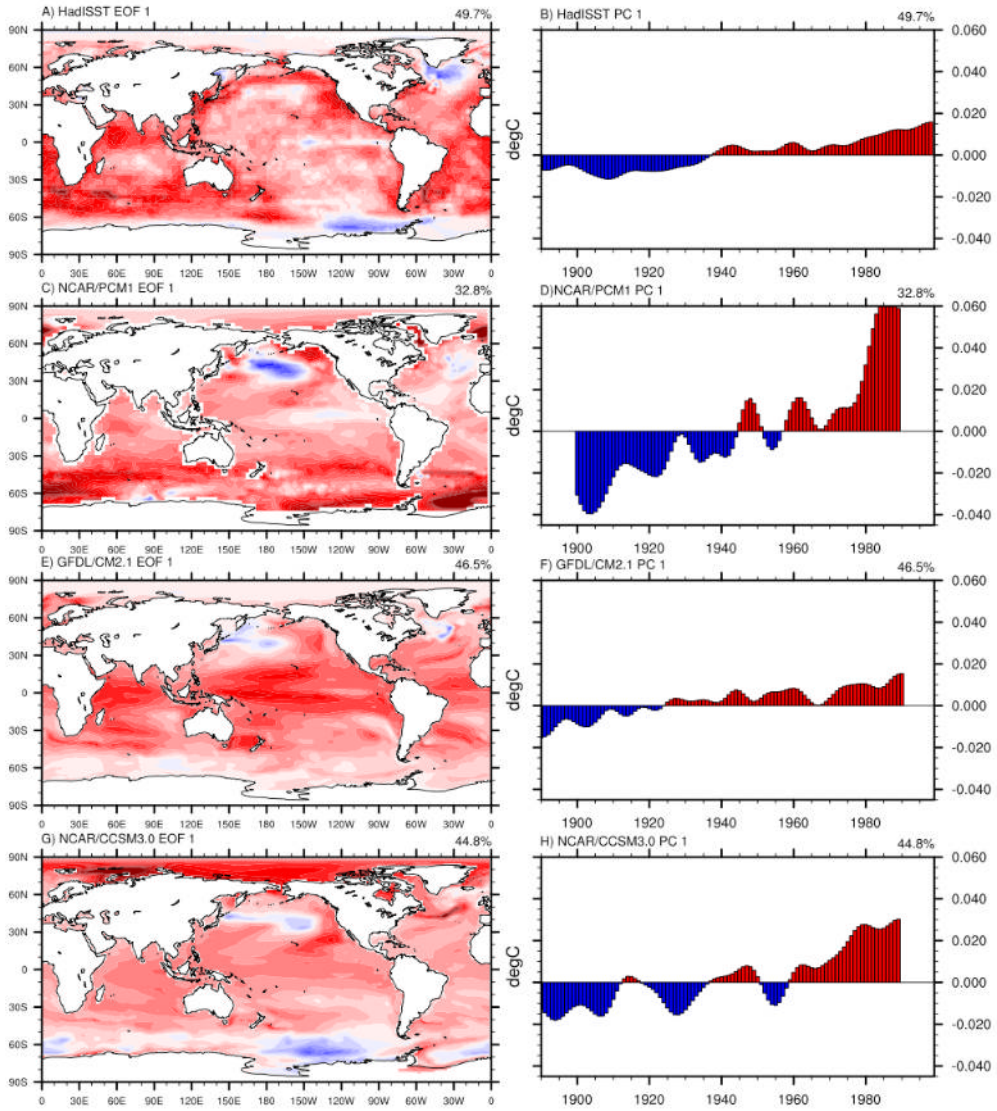


Figure 8. First empirical orthogonal function (EOF) and associated principal component of annual mean sea surface temperature from observations and three 20th Century simulations for years 1890-1999. (A,B) The Hadley Centre sea surface temperature data set (HadISST, Rayner et al., 2003). (C,D) The National Center for Atmospheric Research/Department of Energy Parallel Climate Model Version 1 (NCAR/PCM1, http://www.cgd.ucar.edu/pcm/). (E,F) The Geophysical Fluid Dynamics Laboratory Climate Model version 2.1 (GFDL/CM2.1, Delworth et al. 2006). (G,H) The National Center for Atmospheric Research Community Climate System Model version 3.0 (NCAR/CCSM3.0, http://www.ccsm.ucar.edu/models/ccsm3.0). All data has been smooth with a 10-year low-pass Lanczos filter using 21 weights. EOF patterns are normalized. Principal components are in units of °C. The percent in the upper right of each figure indicates the amount of variance explained by each pattern. Note that the principal component time series from the climate model simulations show fluctuations with larger amplitude than observations, all of which fluctuate on different time scales.