Multidecadal Climate Variability in Observed and Modeled Surface Temperatures*

SERGEY KRAVTSOV AND CHRISTOPHER SPANNAGLE

Department of Mathematical Sciences, Atmospheric Science Group, University of Wisconsin—Milwaukee, Milwaukee, Wisconsin

(Manuscript received 31 January 2007, in final form 8 May 2007)

ABSTRACT

This study identifies interdecadal natural climate variability in global surface temperatures by subtracting, from the observed temperature evolution, multimodel ensemble mean based on the World Climate Research Programme’s (WCRP) Coupled Model Intercomparison Project phase 3 (CMIP3) multimodel dataset. The resulting signal resembles the so-called Atlantic multidecadal oscillation (AMO) and is presumably associated with intrinsic dynamics of the oceanic thermohaline circulation (THC). While certain phases of the oscillation are dominated by the anomalies in the North Atlantic region, other phases are characterized by global teleconnections to the North Pacific Ocean, tropical Atlantic Ocean, as well as the Southern Ocean. In particular, natural variability of sea surface temperature in the Atlantic hurricanes’ main development region has a peak-to-peak amplitude comparable in magnitude to this region’s surface temperature increase over the past century, for all seasons. Evidence suggests that the AMO influence on secular trends in the global-mean surface temperature can arise via direct, regional contribution to the surface temperature evolution, as well as via an indirect route linked to variability of the oceanic uptake of CO₂ associated with AMO-related THC changes.

1. Introduction

Nonuniformity in the global warming trend is usually attributed to corresponding nonuniformities in the external forcing (Crowley 2000). A complementary hypothesis involves multidecadal climate oscillations affecting the rate of global temperature change (Folland et al. 1986; Mann and Park 1994; Schlesinger and Ramankutty 1994); a possible mechanism for this variability is associated with intrinsic dynamics of the oceanic thermohaline circulation (THC; Delworth and Mann 2000; Knight et al. 2005). In the present paper, we use twentieth-century observations of global surface temperature combined with analyses of coupled general circulation model (GCM) simulations in an attempt to differentiate between the externally forced and natural aspects of the observed temperature trends.

Investigators have routinely employed time filtering combined with some procedure to reduce the number of spatial degrees of freedom [typically using empirical orthogonal function (EOF) analysis; Preisendorfer (1988)] in climate change detection–attribution studies (Tett et al. 1996, 1999; Hegerl et al. 1997; North and Stevens 1998; Allen and Tett 1999; Folland et al. 2001). Moron et al. (1998) and Mann and Park (1999) have discussed the application of advanced nonparametric spectral techniques (in which the bandwidth and the shape of filters used to identify signals on various time scales are derived data adaptively, rather than in an ad hoc fashion) to climate data [see Ghil et al. (2002) for the most recent review of these methods]. Moron et al.’s (1998) paper also contains a fairly complete account of observational studies emphasizing surface climate variability. Recently, Schneider and Held (2001) and DelSole (2006) considered nonparametric signal detection methods that use a linear combination of surface temperature data’s leading EOFs to identify robust lowest-frequency spatiotemporal signals.

Our current emphasis is on large-scale secular trends. We will therefore average seasonal surface temperature datasets over 12 subregions of the global domain (see Fig. 1), and filter them in time using a 10-yr boxcar running average. One obvious advantage of this...
procedure compared to more complex methods referenced above is its apparent simplicity. Another advantage is that our processing of observational and GCM-generated data will proceed in absolutely the same way, in spite of the fact that the GCM output is documented on a regular spatiotemporal grid, while the observations are irregular and incomplete. The downside, of course, is the ad hoc nature of filters used, but it seems reasonable to assume that the signal-to-noise ratio in the spatially and time averaged data is high. Note that, in the present paper, the “signal” is conceptually defined as the variability potentially predictable on decadal scales, which includes, in addition to climate changes attributed to a specific forcing, the natural variability characterized by some degree of periodicity in interdecadal range. The “noise,” in contrast, is related to the irregular component of natural variability, which is unpredictable on decadal time scales.

We thus implicitly assume that our spatiotemporal smoothing effectively filters the noise so defined. To separate forced and natural climate signals, we also assume that the grand multimodel ensemble of the twentieth century climate simulations (see section 2b) provides an accurate estimate of the climate response to the time-varying forcing; the difference between the observed variability and the forced response will thus define here the intrinsic climate signal. Note that the GCMs use the “observed” time-varying forcing, so the above estimate of the forced climate response may, in principle, include a contribution related to the natural variability if some of the forcings are affected by long-term natural climate signals. We will discuss this issue further in section 5.

The regions in Fig. 1 are chosen subjectively to represent four latitudinal bands and Pacific, Atlantic, and Indian–Eurasian sectors, respectively. While the majority of the data points within each subregion do belong to the geographical region emphasized in the sector label, there is some degree of arbitrariness in the assignment of these labels; for example, the “South Atlantic (AS)” region contains points within the South Pacific Ocean, and so on. However, our primary interests are in (i) identifying large-scale climate patterns associated with interregion teleconnections, and (ii) establishing the systematic differences between observed and modeled variability within each subregion. Studying either of the above topics does not require specification of exact location and shape of the subregions. We will address these two primary questions in sections 3 and 4a by computing leading EOFs of multiregion observational data (the observational datasets are described in section 2a) and data–model ensemble difference (the model simulations are described in section 2b), respectively.
The results of sections 3 and 4a will point to an intriguing correspondence between the leading EOF of the data–model difference and an EOF pair (EOFs 2 and 3) of observational data that describes a 60–80-yr climate signal most pronounced in the Atlantic sector, hence termed the Atlantic multidecadal oscillation (AMO; Schlesinger and Ramankutty 1994). AMO has been shown to affect regional precipitation characteristics (Enfield et al. 2001; Rogers and Coleman 2003; McCabe et al. 2004; Sutton and Hodson 2005). Goldberg et al. (2001) have speculated that AMO may have contributed to the recent increase in the hurricane activity by adding significantly to tropical Atlantic warming; alternative explanation attributes most of the tropical Atlantic warming to human-induced climate change (Emanuel 2005; Webster et al. 2005; Elsner 2006; Mann and Emanuel 2006; Santer et al. 2006; Trenberth and Shea 2006). Section 4b of the present paper discusses seasonal dependence of interdecadal natural climate variability (defined here, once again, as the residual between smoothed observational data and multimodel ensemble simulations) in the tropical Atlantic Ocean and argues for a larger influence of the AMO than contemplated in the studies that have lobbied for the hypothesis of dominating anthropogenic influence on multidecadal variability of Atlantic hurricanes; our findings are consistent with recent modeling results by Zhang and Delworth (2006), Knight et al. (2006), and Zhang et al. (2007). The summary of our results and the discussion of their implications are presented in section 5. Supplemental materials available at http://dx.doi.org/10.1175/2007JCLI1874.s1 contain further details on the patterns of natural variability identified in the present paper, and repeat our analyses using an alternative surface temperature dataset.

2. Surface temperature data

a. Observational datasets

For our study, we used the Goddard Institute for Space Studies (GISS) analysis of surface air temperature (SAT) station data from meteorological stations (Hansen et al. 1981; Hansen and Lebedeff 1987) combined with sea surface temperature (SST) data, as described in Jones et al. (1999) and Hansen et al. (1999, 2001). The SST measurements are based on ship data prior to 1981 (HadISST1 dataset; Rayner 2000; Rayner et al. 2003, and include satellite observations after that time (Reynolds and Smith 1994; Reynolds et al. 2002; Smith and Reynolds 2004). Combining land and marine temperature data may result in substantial errors in estimation of global surface temperature before 1900, due to poor coverage of ship data; the errors are apparently large enough to completely mute the global cooling after the 1883 Krakatau eruption (Hansen et al. 2005). However, our analyses emphasize longer-time-scale, post-1900 features, so that discrepancies between data and models prior to 1900 are not an issue here.

The GISS temperature data (http://data.giss.nasa.gov/gistemp/) are documented on a nonuniform grid of 8000 equal-area boxes; the temperature estimates within each box were obtained using a spatial smoothing of concurrent observations within a certain distance R of this box. We used, primarily, the dataset for which R = 250 km; the global coverage of surface temperature (i.e., fraction of months with data to the total number of months in 1880–2006) is shown in Fig. 2.1

Our analyses included all grid points between 40°S and 60°N, as well as the channels with data coverage exceeding 0.8 elsewhere. Prior to processing, we filled missing data points in each channel by fitting a principal component regression (PCR; Press et al. 1994) model relating the temperature in a given grid point to that in the N nearest grid points with complete data coverage, as well as to the first harmonic of the seasonal cycle [sin(2πt/12), cos(2πt/12)] (time t is measured in months). The number N equaled to 1/10 the number of available data points for a given grid point. In inverting the singular value decomposition (SVD) of the design matrix to get the regression coefficients, the singular values (or rather, their inverses) with magnitudes below 10% of the leading singular value were neglected. Both the choice of N and the PCR regularization procedure were designed to avoid overfitting.

The subregion (Fig. 1) averages of the surface temperature only included the channels with complete temporal data coverage, after filling in the missing data (see above). In Fig. 2, therefore, the boundaries of the saturated red color (original coverage >80%) poleward of 40°S and 60°N mark the region of observational data channels considered. Comparing this region with the location of the 12 subregions in Fig. 1, we see that the subregions are also characterized by a fairly complete spatial coverage, including the polar subregions.

To assess how robust our results are with respect to pre- and postprocessing data, we repeated our analyses using the alternative surface temperature dataset released jointly by the Hadley Centre and Climatic Research Unit of the University of East Anglia (HadCRUT3; 1 We also repeated our analyses using a version of the GISS dataset with R = 1200 km (not shown) and found that most of the results are recovered, with the exception of artificial interdecadal trends in the tropical South Atlantic in the smoother dataset; these latter trends are, therefore, clearly due to oversmoothing.
Brohan et al. 2006). This set (http://www.hadobs.org and http://www.cru.uea.ac.uk/cru/data/temperature/) updates earlier versions (Jones 1994; Jones and Moberg 2003) to include improvements in the marine component of the temperature data (Rayner et al. 2006). We did not attempt any special treatment of the missing data when processing this dataset. In computing sub-region averages at a given time, we added area-weighted contributions from individual grid points with data available at that time, and divided the result by the total area of all such grid cells; in other words, the average over the available data was assumed to represent the subregion average. Despite differences in the processing of the GISS and HadCRUT3 datasets, they produce consistent diagnosis of secular variability and data–model differences, with some discrepancies on a regional scale (see section 4b and supplemental materials). We have also confirmed the robustness of our results using the extended version of the SST product by Kaplan et al. (1998); to produce this set, the data were optimally interpolated by using spatial coherence information to fill missing data gaps. Both HadCRUT3 and Kaplan dataset versions used here were documented on a 5° × 5° uniform grid, at monthly temporal resolution.

b. GCM-generated datasets

We analyzed simulations of the twentieth-century climate (the so-called 20c3m run) performed by various modeling groups in support of the World Climate Research Programme’s (WCRP) Coupled Model Intercomparison Project phase 3 (CMIP3; see Table 1). Each row of this table lists the identification (ID) of a given model, the corresponding research group–country, and the number of 20c3m realizations analyzed; within a particular model, multiple realizations employed the same variable forcing, but used different initial conditions. We only considered the simulations that provided continuous monthly coverage in the period of 1880–1999. The 20c3m simulations were forced by historical changes in a number of anthropogenic and natural forcing agents, as described on the Web site referenced in the caption of Table 1. Supplemental online material in Santer et al. (2005, 2006) provides table summaries of the forcings used in 20c3m runs (see http://www.sciencemag.
org/cgi/content/full/sci;1114867/DC1 and http://www.pnas.org/cgi/content/full/0602861103/DC1, respectively).

All model simulations include the well-mixed greenhouse gas forcing and the direct effects of sulfate aerosols; otherwise, the differences in applied external forcing across the models are substantial. In particular, only 9 of 16 models listed in Table 1 account for volcanic aerosol forcing and the stratospheric ozone effects, while only 5 of 16 include indirect effects of sulfate aerosols. The uncertainties in the estimates of the forcing are fairly substantial (Crowley 2000; Schwartz 2004); however, all the models reproduce the net twentieth-century’s global temperature increase quite well. We therefore interpret the grand ensemble average over 52 simulations listed in Table 1 as our best available estimate for the externally forced surface temperature variability,2 since subdecadal variability is eliminated by our time filtering, while longer-term, decadal and longer time-scale natural variability present are largely smoothed out in the ensemble averaging. The spread of the individual-model ensemble means provides an estimate of the forced-signal uncertainty in the models, which combines the uncertainties due to structural model differences and those due to different external forcing used.

c. Filtering and averaging procedures

We first removed the seasonal cycle from the monthly surface temperature data by retaining only the residual of the linear regression of each grid point’s time series onto the first five harmonics of the seasonal cycle. The El Niño–Southern Oscillation (ENSO) signal was removed in the same way by linear subtraction of the variability correlated with Niño-3 index time series. Next, we formed seasonal averages [December–February (DJF), March–May (MAM), June–August (JJA), and September–November (SON)] of the dataset so obtained, and applied to them a 10-yr boxcar running filter to concentrate on multidecadal time scales. Finally, the subregion averaging was performed, providing us with a multiregion time series of secular climate trends in the past century for each season.

The procedure above does not depend on the order of averaging operations as long as the fixed spatial distribution of continuously sampled data channels is available, which is the case for the postprocessed GISS dataset and model output data described in sections 2a, b. In the case of irregular and incomplete data, such as the HadCRUT3 dataset discussed in more detail in supplemental materials, we first applied spatial averaging to get unfiltered subregion anomalies, which was followed by seasonal cycle and ENSO signal subtraction, seasonal averaging, and decadal smoothing.

Finally, when comparing the various datasets, we subtracted from the surface temperature time series at each grid point the temperature’s average value in the period 1951–80; the time series presented in section 4 are the anomalies relative to this base period.

3. Secular variability in the GISS dataset

The time series [principal components (PCs)] of the leading EOFs associated with the centered and normal

---

2 This, once again, assumes that the variable forcing used in the models does not have a component due to natural variability, which may not be the case (see section 5).
ized, multiregion (Fig. 1), annual surface temperature dataset (GISS analysis) are shown in Fig. 3. The leading PC represents a nonuniform global warming trend (characterized by highest cross-region temporal correlations); the next two time series, as well as the associated spatial patterns (not shown here; see supplemental materials), are in quadrature and strongly hint at a multidecadal oscillation with the time scale of 60–80 yr. The latter patterns were obtained by regressing surface temperature upon the time series of PC-2 and PC-3. Multiplying these patterns by the values of corresponding PCs at a given time and adding the results produced the reconstruction of the multidecadal oscillation displayed in Fig. 4.

In certain phases (e.g., see panels of Fig. 4 describing years 1929, 1979), this oscillation is dominated by the seesaw-type anomalies in the North and South Atlantic Ocean; these anomalies are most pronounced in climatically important regions of deep-water formation in the Labrador and Weddell Seas. At other times (e.g., in years 1939 and 1989), the anomalies are confined to the North Atlantic only. Finally, some phases are characterized by substantial anomalies in the midlatitude North Pacific Ocean, which are also accompanied by opposite-sign anomalies northward spreading across the whole latitudinal circle (e.g., years 1949–69).

The 1929-type phase closely resembles, in its time scale, magnitude, and spatial pattern, the AMO signal of Knight et al. (2005, their Fig. 1), although our plots cover a slightly larger portion of the globe than the 40°S – 65°N belt displayed by these authors. The global character of the anomalies and dominance of the anomalies in the deep-water formation region in Fig. 4 are both consistent with an inferred dynamical connection between the AMO and natural variability of THC (Delworth and Mann 2000; Knight et al. 2005). We speculate here that the SST anomalies in the midlatitude Pacific Ocean during certain phases of the oscillation are due to local heat-flux forcing associated with the atmospheric jet’s displacements in response to the ocean-induced SST anomalies in the Atlantic; similar speculation appeared in Enfield et al. (2001).

The multidecadal AMO-type variability described above has a much smaller amplitude than that of the global warming trend (see Fig. 3), and a very weak projection on the global-mean temperature time series, which is consistent with findings of Knight et al. (2005). In contrast, Zhang et al. (2007) have argued for a larger direct contribution of the multidecadal Atlantic SST variability to the global temperature signal. We

---

3 Given our procedure, this fact becomes trivial after noting that the leading PC in Fig. 3 is well correlated with the global temperature time series (see Fig. 5).
will also speculate in section 5 on a possibility of indirect AMO contribution to global temperature variability. The basis for this speculation is an observation that the leading “global warming” PC exhibits local troughs at about 1910 and 1970, as well as a crest at 1940, thus suggesting a time scale in the global temperature time series consistent with the AMO time scale. Furthermore, there is a substantial overlap in the spatial pat-
tern of the global warming trend in the Northern Hemisphere and the AMO’s spatial pattern (see supplemental materials). Finally, the analysis of section 4 argues that the interdecadal wiggle in the global temperature time series cannot be explained by the variability in the external forcing. An emerging hypothesis is, therefore, that the AMO variability may still contribute to the globally averaged surface temperature variability, although not directly, but via affecting concentrations of well-mixed greenhouse gases in the atmosphere on a decadal time scale (see section 5).

In summary, secular variability in the observed surface temperature is a combination of a nonuniform global warming trend and a multidecadal oscillation similar to the AMO signal reported elsewhere (Delworth and Mann 2000; Knight et al. 2005). Note, however, that we did not employ a much criticized linear detrending of surface temperature data (Mann and Emanuel 2006; Trenberth and Shea 2006) to define the AMO index [D. B. Enfield and L. Cid-Serrano (2007, personal communication) defined the AMO index by subtracting a quadratic fit from the North Atlantic surface temperature time series]. Note also that our global warming PC-1 and multidecadal PC-3 are very similar to the time series of the leading optimal persistence patterns (DelSole 2001) of the global surface temperature data reported by DelSole (2006).

4. Comparison of model simulations with observations

a. Results using annual-mean data

We first computed, for each of the 16 models listed in Table 1, the single ensemble mean surface temperature time series by averaging over all available runs of this model [the Geophysical Fluid Dynamics Laboratory (GFDL) model versions 2.0 and 2.1 were combined to form a single ensemble mean]. These time series were then averaged to get the multimodel ensemble mean temperature time series. The results for globally averaged temperature are presented in Fig. 5, which shows the multimodel ensemble mean anomaly relative to the 1950–81 base period (black line) along with the ensemble standard deviation (dashed lines), superimposed on the observed global temperature anomaly (red line). Taking into account uncertainties associated with the observed anomaly (not shown), we conclude that the models do a good job in representing global temperature variability. However, local troughs in the 1910s and 1970s and a crest around 1940 constitute notable structural deviations of the observed anomaly relative to its simulated counterparts. That is to say that while individual models may produce anomalies of proper sign and magnitude in 1910, 1940, and 1970 (as documented by the dashed lines in Fig. 5), not a single
model captures the observed succession of all three extreme events. We interpret these events as a manifestation of a multidecadal natural signal (associated with the AMO) superimposed onto a human-induced warming trend.

To do so, we compared observed and simulated regional anomalies associated with subregions of Fig. 1. The results are shown in Fig. 6, which has a layout analogous to that of Fig. 5, but the spaghetti plots of individual-model members of the grand multimodel ensemble are used in place of standard deviation error bars. The subregion labels are given in the heading of each panel; for viewing convenience, the location of the panels corresponds to relative location of subregions in the map of Fig. 1 (Pacific region is leftmost, North Pole is toward the top of the figure).

The structure of data–model differences in the Northern Hemisphere has a character similar to that of globally averaged temperature differences: it is dominated by an oscillatory signal of positive sign in the 1940s and negative sign in the 1910s and 1970s; this signal is most pronounced in the Pacific and Atlantic sectors, including the tropical Atlantic (we will come back to this important issue in section 4b). The models closely track, in general, the observed Southern Hemisphere’s surface temperatures in the second half of the twentieth century and show considerable deviations from data-derived trends before that, especially in the Southern Ocean region. The poor coverage and quality of the early twentieth century data in the latter region, however, preclude any definitive conclusions about the exact nature of data–model differences there. Therefore, we have excluded the PS, AS, and IS regions (in Fig. 1) from the further EOF analysis in the present subsection.4

We isolated the leading mode of model simulations’ discrepancies from the observed surface temperature variability by performing EOF analysis on 9-region (i.e., excluding the 3 Southern Ocean regions from the 12 original subregions) time series of data–model differences. The anomaly of the latter difference for each subregion was weighted by the square root of this region’s area, which ranks the EOF modes in terms of their contribution to the area-averaged variance. The resulting time series (Fig. 7) is very well correlated with the AMO index defined via linear detrending of the North Atlantic SSTs (Enfield et al. 2001; McCabe et al. 2004; Sutton and Hodson 2005; Knight et al. 2005). Note, however, that no linear detrending has been applied here. Furthermore, the AMO index of Fig. 7 presents a measure of natural climate variability free of statistical assumptions used in Mann and Emanuel (2006) and Trenberth and Shea (2006) to separate the influence of the externally forced climate trend.

We interpret the new AMO index defined above as a measure of the multidecadal natural climate variability. This interpretation is subject to our assumption that the multimodel ensemble average response can be taken as the forced signal for the actual observations. The GCM members of this ensemble are prone to parameterization errors, as well as uncertainties in the external forcings. We implicitly assume that these model errors are not systematic, so that the ensemble averaging across all the models will reduce both types of uncertainties. To check whether the AMO multidecadal variability is statistically significant in the background of intermodel differences associated with forcing and parameterization uncertainties, we repeat the model–data difference EOF analysis above for 1000 surrogate 15-model ensembles obtained via bootstrap resampling, by picking each set randomly from the collection of the available models with replacement.5 We then compute the leading PC of each surrogate set and sort, for each year, the 1000 surrogate PCs’ values in ascending order. The dashed lines in Fig. 7 show the time series of the 25th and 975th sorted PC values so obtained. Note that the negative values of the AMO index in the 1920s and 1980s, as well as its positive values during 1940–70, are statistically significant at the 95% level.

The spatial pattern associated with the AMO index of Fig. 7 can be obtained by regressing gridded data–model difference time series onto the normalized AMO time series (see Fig. 8). This pattern strongly resembles the one associated with a 1929–39 phase of the observed multidecadal signal discussed in section 3, which in turn agrees well with the AMO pattern of Knight et al. (2005), thus underscoring the interpretation of this mode as the one associated with the natural climate variability put forth in section 3.

To summarize, both the time scale and pattern of the leading mode of data–model differences described in this section are consistent with the multidecadal AMO signal of section 3. As the multimodel ensemble mean represents an estimate of the externally forced climate signal, the AMO appears to be due to intrinsic climate dynamics, in agreement with previous studies.

4 The multiregion EOF results of section 3 are not sensitive to this exclusion.

5 In this way, each surrogate set may contain identical members, while missing some of the model runs.
Fig. 6. Observed and simulated regional surface temperature anomalies averaged within 12 subregions shown in Fig. 1. Red lines show the observed temperature anomaly (relative to 1951–80 base period), while heavy black lines represent the multimodel ensemble mean. Surface temperature time series associated with individual models' ensembles are plotted as light colored lines. The heading of each panel indicates the subregion being considered.
b. Seasonal dependence and tropical Atlantic AMO influence

The results of Figs. 6 and 8 argue for a substantial annual-mean AMO influence on the tropical Atlantic SSTs. In particular, the TAN panel of Fig. 6 shows considerable deviations of the total observed surface temperature signal from the model estimate of the externally forced component. Furthermore, an inherent part of the AMO pattern in Fig. 8 is a tongue of warm SST anomalies in the tropical Atlantic. The issue of the AMO influence on the tropical Atlantic SSTs is an im-

Fig. 7. Time series of the leading EOF of multiregion data – model surface temperature anomaly difference. The dashed lines are 2.5 and 97.5 percentiles of the errors associated with the AMO index so defined. These significance estimates were computed using bootstrap resampling of the members of the multimodel ensemble (see text for details). The negative values of the AMO index in the 1920s and 1980s, as well as its positive values during 1940–70, are statistically significant at the 95% level.

Fig. 8. The spatial pattern associated with the time series of Fig. 7.
portant one because of a possible influence on the tropical cyclogenesis (see the corresponding discussion in section 1). While the above results imply a connection, they were obtained using large-scale spatial and annual time averaging, whereas tropical cyclone generation occurs on a smaller regional scale and is season-dependent.

Because of an utmost importance of the implied AMO influence on hurricanes, we now discuss the seasonal dependence of our results using an example of the tropical Atlantic SSTs. The time series of observed and multimodel-simulated seasonal SST anomalies averaged over the tropical cyclone main development region (6°–18°N, 60°–20°W) is shown in Fig. 9, each panel of which has the same layout as in Fig. 6. We see that significant AMO-related SST anomalies (defined here, once again, as differences between the observations and grand multimodel ensemble) arise for all seasons, including the season most relevant to tropical cyclone formation (August–October); these anomalies have a magnitude not captured by a majority of individual model members and temporal structure captured by none of these models. The peak-to-peak amplitude of SST variations associated with natural climate signal is on the order of externally forced temperature rise in the period of 1900–2000.

The same computation as above using the Kaplan SST dataset instead of GISS dataset of Fig. 9 produces essentially the same results (see supplemental materials), while the results of using the HadCRUT3 dataset indicate a smaller but still significant AMO influence on the tropical SSTs (Fig. 10).

The above results thus argue for a sizeable influence of the AMO-type natural variability onto the tropical North Atlantic SSTs throughout the year, with possible significant implications for hurricane formation in this region. The evidence presented here places this tropical AMO influence in the middle range in relation

![Fig. 9](image-url)
to studies that emphasize either the natural variability (Goldenemberg et al. 2001) or external anthropogenic forcing (Emanuel 2005; Webster et al. 2005; Elsner 2006; Mann and Emanuel 2006; Santer et al. 2006; Trenberth and Shea 2006). Our results are also consistent with recent modeling studies by Zhang and Delworth (2006), Knight et al. (2006), and Zhang et al. (2007), as well as with the analysis of D. B. Enfield and L. Cid-Serrano (2007, personal communication).

5. Summary and discussion

In this paper, we examined secular climate variability in observed and simulated surface temperatures of the past century. We used spatial (Fig. 1) and temporal decadal smoothing to increase the signal-to-noise ratio. The grand multimodel ensemble mean of the twentieth-century simulations documented in the WCRP CMIP3 multimodel dataset (see Table 1) provided an estimate of the forced climate response, as ensemble averaging presumably reduced errors associated with both the models’ physics and different variable forcings used by the models; we interpreted the residual observed variability as the natural climate signal.

Multiregion EOF analysis of the observed surface temperatures (section 3; Fig. 3) identified a nonuniform global warming trend, as well as a 60–80-yr oscillation whose spatial pattern (Fig. 4) resembles the so-called Atlantic multidecadal oscillation (AMO) reported elsewhere (Folland et al. 1986; Mann and Park 1994; Schlesinger and Ramankutty 1994; Delworth and Mann 2000; Enfield et al. 2001; Sutton and Hodson 2005). We showed in section 4a that the data–model differences (Figs. 5, 6) are dominated by a signal of a similar time scale and spatial pattern as the AMO (Figs. 7, 8, respectively), providing further support for attributing this signal to natural climate variability associated with the oceanic thermohaline circulation (THC; Delworth et al. 1993; Delworth and Mann 2000; Knight et al. 2005).

Fig. 10. Same as in Fig. 9, but using the HadCRUT3 dataset.
Section 4b dealt with the seasonal dependence of the AMO (defined as the difference between observations and grand multimodel ensemble) and concentrated on the SST variability in the tropical Atlantic Ocean—the region in which cyclogenesis occurs during the boreal summer–fall season. The results (Fig. 9) argue for a larger influence of the AMO on the tropical Atlantic SSTs than contemplated in some previous studies (Emanuel 2005; Webster et al. 2005; Mann and Emanuel 2006; Santer et al. 2006; Trenberth and Shea 2006), although not quite as pronounced as suggested by Goldenberg et al. (2001); they are consistent with model-based conjectures of Zhang and Delworth (2006), Knight et al. (2006), and Zhang et al. (2007), as well as with the analysis of D. B. Enfield and L. Cid-Serrano (2007, personal communication). The degree of tropical SST association with the AMO is somewhat sensitive to the observed SST dataset used (Fig. 10; also see supplemental materials).

The main assumption of our analysis is that the ensemble mean calculated from the WCRP CMIP3 multimodel dataset provides an accurate estimate of the climate response to time-varying forcing. Many of the CMIP3 models have alike biases in the simulated climatology of tropical Pacific, tropical Atlantic, and mid-latitude North Atlantic SSTs. In an analogous way, the estimate of the climate response to time varying forcing may also have similar biases across these models (see, e.g., the pre-1950 period of the TIS panel of Fig. 6). On the other hand, observed surface temperature by itself is prone to observational and processing errors (most notably at the Southern Ocean region in the beginning of the twentieth century). In both of the latter cases, the difference between the multimodel ensemble mean and the observed surface temperatures will in fact provide an estimate of the natural climate variability contaminated by the aforementioned model biases. We addressed the issue of uncertainties associated with model and forcing biases by performing bootstrap resampling of the intermodel ensemble members and showed that the AMO index defined as the leading PC of multimodel ensemble mean is characterized by statistically significant deviations from zero throughout most of the twentieth century (see Fig. 7). Our analysis confirms earlier conclusions about the AMO teleconnections—in particular, the links between the North Atlantic and North Pacific (Enfield et al. 2001; Sutton and Hodson 2005), as well as with the tropics (Goldenberg et al. 2001). However, the cycle shown in Fig. 4 suggests a truly global character of the AMO teleconnections, including concurrent temperature anomalies in the regions of deep-water formation in the Northern and Southern Hemispheres. While this particular result is inconclusive, due to the Southern Hemisphere data’s poor quality and coverage in the beginning of the twentieth century, it is conceptually consistent with the inferred connection between the AMO and the Atlantic THC (Delworth et al. 1993; Delworth and Mann 2000; Knight et al. 2005). We speculate that the AMO-related SST signatures in the Pacific occur via an atmospheric route, and are due to local atmospheric forcing associated with the jet stream response to the ocean-induced SST anomalies in the Atlantic Ocean [a similar argument was presented by Enfield et al. (2001)].

Traditional interpretation of the decreasing global-mean surface temperature during the period of 1940–70 is that the tropospheric aerosols’ cooling effect, possibly combined with solar forcing variability, outweighed (greenhouse gas) GHG-induced warming during this period (Crowley 2000; Meehl et al. 2003). Both the tropical aerosol and GHG forcings are incorporated in the model runs we have analyzed; yet, the simulated global temperature time series underestimates this local cooling trend (Fig. 5). Previous studies have associated the anomalous (i.e., not captured by most of the models) warmth in 1940 with the natural climate variability (Delworth and Knutson 2000; Johannessen et al. 2004). We argue here specifically, however, that this signal is a part of the AMO, as the discrepancies between the observed and simulated surface temperature variability occur on a multidecadal time scale both prior to and after 1940, and encompass multiple regions throughout the globe (Fig. 6).

Perhaps the most important and also the most controversial question is that of the AMO influence on the tropical Atlantic SST. Mann and Emanuel (2006) constructed a skillful linear regression model of the tropical Atlantic SST evolution using the globally averaged surface temperature and tropospheric aerosol forcing time series (Crowley 2000) as predictors to argue for a negligible AMO influence. The motivation to include the second predictor comes from modeling studies showing a significant effect of the aerosol forcing on the SST in the region of interest (Hansen et al. 2005). The estimates of this forcing are, however, highly uncertain (Schwartz 2004); these uncertainties are likely to undermine the conclusions solely based on the regression analysis.

Furthermore, the global temperature signal in Mann and Emanuel’s study is interpreted as being entirely due to external forcing. This assumption effectively leads to a modified AMO index (Trenberth and Shea 2006), which is less correlated with tropical Atlantic SSTs than the one obtained by linear detrending of
the North Atlantic temperatures (Enfield et al. 2001; McCabe et al. 2004; Sutton and Hodson 2005; Knight et al. 2005). Our present analysis made an alternative assumption that the grand multimodel ensemble mean provides the best estimate of externally forced signal available (which is neither perfectly correlated with global temperature nor linear) and defined the revised AMO index as the leading PC of the multiregion data–model difference [D. B. Enfield and L. Cid-Serrano (2007, personal communication) provide yet another alternative definition of the AMO index]. Quite surprisingly, our new AMO index corresponds almost perfectly to the one based on linearly detrended North Atlantic SSTs, consistent with a stronger connection to the tropical Atlantic compared to the one implied by Trenberth and Shea (2006) and Mann and Emanuel (2006).

Finally, Santer et al. (2006) show that the level of intrinsic variability in the current generation of climate models is insufficient to explain, by itself, the observed interdecadal evolution of the tropical Atlantic SSTs during the past century [see, however, Zhang and Delworth (2006), Knight et al. (2006), and Zhang et al. (2007) for contrary examples], and attribute a dominant portion of 1900–2000 tropical Atlantic warming to the externally forced signal. These results are really in no contradiction, qualitatively and quantitatively, with those presented here, but the thrust of interpretation in the two papers is fundamentally different. In particular, we think that the influence of multidecadal natural climate variability onto the tropical Atlantic climate should be taken into consideration to improve our projections of future climate change, while Santer et al. (2006) seem to suggest that it is safe to fully neglect this influence.

One of the reviewers pointed out an interesting difference in seasonal dependence between the observed and modeled tropical SST evolution. Namely, the observations from GISS, HadCRUT3, and Kaplan datasets all show substantial differences in the tropical North Atlantic SST between winter–spring and summer–fall seasons during the period of 1980–2000 (see Figs. 9, 10 here, as well as Fig. 9 of the supplemental materials), that is, less warming in winter–spring and more warming in summer–fall seasons. On the other hand, the simulated tropical North Atlantic SST systematically shows very little change between different seasons. Two possibilities for such a discrepancy are as follows: (i) this seasonal dependence is a characteristic feature of the AMO variability; and (ii) climate GCMs fail to capture the seasonal response to the external radiative forcing during the period of 1980–2000. We plan to address this issue in future work.

The spatial pattern of the inferred multidecadal signal in Fig. 4 has a small global average, which is consistent with a fairly small direct contribution of the AMO variability to global temperature evolution, in line with the study of Knight et al. (2005). However, a recent modeling work (Zhang et al. 2007) provided evidence to the contrary of this statement. The latter authors simulated the impact of AMO on the multidecadal variability of Northern Hemisphere mean surface temperature and found that the AMO-like oceanic fluctuations could have added substantially to the early twentieth-century warming, the midcentury pause in warming, and the rapid warming of the past few decades.

In the present paper, the global-mean temperature trend is shown to be almost entirely accounted for by the leading PC of the multiregion data (Fig. 3); the latter PC, when linearly detrended, is a temperature anomaly whose apparent time scale is consistent with the multidecadal signal of PCs 2 and 3. An intriguing correspondence, in terms of time scales (as well as in terms of the spatial pattern in the Northern Hemisphere; see supplemental materials), between the global and AMO-related temperature anomalies prompts an interesting question, which we plan to address further in future work, namely, “Is this global temperature anomaly due, in part, to changes in the ocean uptake of CO2 associated with multidecadal natural signal (AMO)?”

Figure 11 shows linearly detrended atmospheric CO2 measurements collected at Mauna Loa Observatory (Keeling et al. 1976; Thoning et al. 1989). Linear trend actually dominates changes in the carbon dioxide concentration during 1959–2006; however, the interdecadal deviations from this trend, while smaller, exhibit an interesting in-phase relationship with the time series of global temperature (Fig. 3, upper panel) and multidecadal variability (Fig. 7). Knight et al. (2005) argue, in a modeling study, that the North Atlantic SST index (leading, in general, our modified AMO index of Fig. 7 by a few years) is a good proxy for the strength of the Atlantic THC. Thus we infer that reduced concentrations of carbon dioxide in 1970–80 (Fig. 11) are associated with the minimum of THC, while positive anomalies in the 1960s and 2000–10 are associated with an intensified THC.

While detecting and attributing CO2 variability is a challenging task (Sarmiento and Gruber 2002), one can speculate that the ocean’s CO2 uptake intensifies as the enhanced THC brings deep water depleted in CO2 to the surface. Back-of-the-envelope calculation for the advective time scale associated with this process gives an estimate of $\tau = H/w \approx 1000 \text{ m}/10^{-6} \text{ m} \cdot \text{s}^{-1} = 10^9 \text{ s}$.
30 yr, where $H$ is a thermocline depth and $w$ is a typical vertical velocity associated with the western boundary THC anomalies. This estimate is consistent with the multidecadal lag between the inferred maximum of THC in the 1940s and minimum of CO$_2$ deviation from the linear trend in the 1970s (Fig. 11). One of the reviewers commented that alternative mechanisms are also possible, one of which is the solubility pump of CO$_2$. In this mechanism, the warmer (colder) North Atlantic surface ocean during the positive (negative) AMO phase leads to lower (higher) solubility of CO$_2$, thus more (less) CO$_2$ is released into the atmosphere. This mechanism argues for an in-phase relationship between THC and atmospheric CO$_2$, and is also consistent with the observations.

Attributing a fraction of long-term CO$_2$ trends to multidecadal natural variability may have important consequences for future projections of climate change. Business-as-usual scenarios estimate the rate of CO$_2$ increase based on current measurements. However, over/underestimation of this rate by as much as 30%–50% is possible solely because of neglecting the potential contribution from the natural climate signals (see caption to Fig. 11), transferring to comparable relative errors in the projected temperature increases. Our results thus emphasize the need for a careful assessment and understanding of both direct and indirect effects of natural variability in an ongoing climate change; a faithful simulation of such effects in numerical climate models will benefit the accuracy of climate forecasts these models produce.

Acknowledgments. We acknowledge the modeling groups for making their simulations available for analysis, the Program for Climate Model Diagnosis and Intercomparison (PCMDI) for collecting and archiving the CMIP3 model output, and the WCRP’s Working Group on Coupled Modelling (WGCM) for organizing the model data analysis activity. The WCRP CMIP3 multimodel dataset is supported by the Office of Science, U.S. Department of Energy. It is our pleasure to thank A. A. Tsonis for encouragement and helpful discussions. Preliminary results from this research were presented at the CLIVAR Workshop on Multidecadal to Centennial Global Climate Variability, Honolulu, HI, November 14–18, 2006. We thank H. Dijkstra, F.-F. Jin, F. Primeau, and A. Timmermann for organizing this exciting and fruitful scientific event. We have benefited much from the workshop discussions, and are particularly grateful to T. Delworth, D. Enfield, J. Knight, M. Mann, N. Mantua, and D. Vyushin for their constructive comments. Some of the independent results presented at the workshop by D. Enfield and J. Knight in relation to the observed and simulated surface temperature variability were analogous to those described in the present paper. We are also thankful to R. Cohen for his feedback on an earlier version of this paper. The present manuscript has been significantly improved by incorporating additional comments and suggestions offered by D. Enfield, R. Zhang, and an anonymous reviewer. Our research was supported by the Department of Energy Grant DE-FG-03-01ER63260 and by the National Aeronautics and Space Administration Grant NNG-06 AG66G-1.

REFERENCES


Rogers, J. C., and J. S. M. Coleman, 2003: Interactions between


