Global Sea Surface Temperature Analyses: Multiple Problems and Their Implications for Climate Analysis, Modeling, and Reanalysis

James W. Hurrell and Kevin E. Trenberth

National Center for Atmospheric Research,* Boulder, Colorado

ABSTRACT

A comprehensive comparison is made among four sea surface temperature (SST) datasets: the optimum interpolation (OI) and the empirical orthogonal function reconstructed SST analyses from the National Centers for Environmental Prediction (NCEP), the Global Sea-Ice and SST dataset (GISST, version 2.3b) from the United Kingdom Meteorological Office, and the optimal smoothing SST analysis from the Lamont-Doherty Earth Observatory (LDEO). Significant differences exist between the GISST and NCEP 1961–90 SST climatologies, especially in the marginal sea-ice zones and in regions of important small-scale features, such as the Gulf Stream, which are better resolved by the NCEP product. Significant differences also exist in the SST anomalies that relate strongly to the number of in situ observations available. In recent years, correlations between monthly anomalies are less than 0.75 south of about 10°N and are lower still over the southern oceans and parts of the tropical Pacific where root-mean-square differences exceed 0.6°C.

While adequate for many purposes, the SST datasets all contain problems of one sort or another. Noise is evident in the GISST data and realistic temporal persistence of SST anomalies after 1981 is lacking. Trends in recent years are quite different between the GISST and NCEP analyses, and this can be partially traced to differences in the processing of in situ data and an increasing cold bias in the NCEP OI data arising from incompletely corrected satellite data. Significant discrepancies also exist in centennial trends from the LDEO and GISST datasets, and these likely reflect the separate treatment of the very low frequency signal in the GISST analysis and questionable assumptions about the stationarity of statistics in the LDEO method.

Ensembles of integrations with an atmospheric general circulation model (AGCM) are used with three of the SST datasets as lower boundary conditions to show that the differences among them imply physically important differences in the atmospheric circulation. Over the Tropics, where masking by internal atmospheric variability is small, SST differences affect moist convection and systematically produce strong responses in the local divergent circulation. A case study shows that analyzed SST differences in the tropical Pacific can be as large as for a moderate El Niño. Such large discrepancies induce local rainfall anomalies up to 8 mm day^{-1} and, in addition to the tropical circulation anomalies, are associated with global teleconnections that influence temperatures and precipitation around the world. Results also show the limitations to using AGCMs when forced by specified SSTs.

The likely sources of the problems evident in the different SST products are identified and discussed. Several of the problems are being addressed by current efforts to reprocess the SST data, which is strongly recommended, but remaining problems demand further attention and attempts to resolve them should continue. The choice among SST analyses used for AGCM simulations, for the atmospheric reanalysis projects, for identifying climate signals, and for monitoring climate is important, as known flaws in the global analyses can compromise the results.

1. Introduction

Perhaps the most important field in climate system modeling is sea surface temperature (SST). Flaws in simulating SSTs are often corrected in coupled atmosphere–ocean model runs through a "flux correction," which adjusts the heat (and moisture) fluxes

between the atmosphere and ocean so that realistic SSTs are reproduced. For atmospheric general circulation model (AGCM) simulations a sequence of SSTs are specified as the lower boundary condition, implying an infinite heat capacity. The imposed SSTs are typically changed in a realistic fashion by specifying either a mean climatological annual cycle or the observed SSTs over some period of time. In the Atmospheric Modeling Intercomparison Program (AMIP) (Gates 1992), for example, observed SSTs are specified beginning in 1979. Knowledge of SSTs as well as sea ice is also required in analyses of atmospheric fields; thus, the integrity of the global reanalyses, for example by the National Centers for Environmental Prediction (NCEP) and the European Centre for Medium-Range Weather Forecasts (ECMWF), depends critically on the specified lower boundary conditions. Reanalyses have been or will be performed from about 1948 to the present.

The need for correct SSTs in coupled simulations stems directly from several feedback processes that would be seriously distorted by inaccurate surface temperatures. Errors at high latitudes, for instance, can greatly impact the sea-ice extent, resulting in too much or too little, with resultant ice-albedo feedbacks potentially exacerbating the original SST errors. Errors in tropical SSTs can greatly impact moist convection and the hydrological cycle, thereby affecting the water vapor feedback and global teleconnections such as those observed during El Niño events.

Given the critical need for correct SSTs, it is important to document and understand how well SSTs are known and whether or not the errors matter. These are the key questions addressed in this paper. We show that, while new methods of interpolating and extrapolating into areas devoid of observations likely have improved the historical SST analyses, important flaws remain in the global SST fields that are used for driving and validating models and in monitoring climate. Moreover, some of these are sufficiently serious that they compromise the results of model simulations and analyses produced with four-dimensional data assimilation (4DDA). Some problems with temporal continuity can be partially ameliorated with smoothing. However, other flaws or uncertainties, such as in trends, are not easily reduced.

We have first performed comprehensive comparisons among four monthly SST datasets: from NCEP both the optimal interpolation (OI) SST analysis of Reynolds and Smith (1994) and the empirical orthogonal function (EOF) reconstructed SST analysis of

Smith et al. (1996), from the United Kingdom Meteorological Office (UKMO) version 2.3b of the Global Sea-Ice and SST dataset (GISST) of Rayner et al. (1996), and from the Lamont-Doherty Earth Observatory (LDEO) the optimal smoothing (OS) SST analysis of Kaplan et al. (1997, 1998). As well as significant differences between the long-term mean climatologies of the NCEP and GISST products, large differences exist in the monthly time series and are dramatically revealed by the autocorrelations of each product. The different SST datasets have been used as lower boundary conditions for ensembles of runs with the National Center for Atmospheric Research (NCAR) GCM (version 3 of the Community Climate Model, CCM3), which enables us to assess the climatic significance of the analyzed SST differences. As we will show, displaced convection in the tropical Pacific because of errors in SSTs has a strong and direct impact on the tropical circulation, as well as on teleconnections into midlatitudes and elsewhere.

Another key issue is the long-term trends in the datasets. New methods for analyzing SSTs use a recent well-observed base period to define spatial structures (modes) and statistics, such as how much a given observation projects onto each mode, that are then used to interpolate in space and time. This procedure has the advantage of more reliably projecting the SST anomaly patterns that exist based on limited observations, but it depends critically on the assumption of stationarity of the statistics. In particular, the presence of trends, such as those expected with climate change, seriously violates the assumptions of stationarity. For instance, given the presence of a linear trend in a long record, the standard deviation of that trend in a subsample of the record is proportional to the length of the subsample. Statistics based on the subsample will, therefore, necessarily underestimate the component that projects onto the long-term trend. Different treatments of the very low frequency signal can give greatly different results. Both the GISST and LDEO datasets extend over a century and feature quite different trends. The LDEO dataset, for example, indicates less warming than the GISST dataset at most locations, including a cooling in the tropical eastern Pacific since the turn of the century (Cane et al. 1997).

Other recent comparisons of SST datasets have been carried out by Trenberth et al. (1992, henceforth TCH) and Folland et al. (1993). A more complete summary of the results of TCH on errors in SSTs and their origins is given is section 6 along with a discussion of the problems in defining SST and the impacts of sat-

ellite retrievals. TCH compared the reproducibility of SSTs in analyses from the UKMO and the U.S. Climate Prediction Center (CPC), which revealed monthly anomaly correlations on a 5° grid exceeding 0.9 over the northern oceans but less than 0.6 in the central tropical Pacific and south of about 35°S. Root-mean-square differences between CPC and UKMO monthly SST anomalies exceed 0.6°C in the regions where the correlation is lower than about 0.6. Similar results are reported here, and the dependence on the number of in situ observations is clear.

The datasets compared and evaluated are described in section 2, and in section 3 the model experiments used to assess the importance of the discrepancies in the SST fields are outlined. The results of the comparisons are presented in section 4, and the results of the model experiments are given in section 5. The results are discussed in the context of other sources of information relating to the reasons why discrepancies exist in section 6, and conclusions are drawn in section 7.

2. SST datasets

The four SST datasets examined here are all monthly but differ in terms of spatial resolution, coverage, and length of record. Each dataset is briefly summarized below, as are two high-resolution SST climatologies. All comparisons between datasets are made over common periods of time and on common grids. The latter sometimes required degrading higher-resolution SST analyses to lower resolutions by simple averaging techniques to minimize aliasing. Trenberth and Solomon (1993) provide a discussion of the errors caused by interpolating from finer to coarser grids.

a. NCEP OI analyses

The OI SST analysis technique described by Reynolds and Smith (1994) was developed for operational purposes at NCEP. It follows on the analysis methods of Reynolds (1988) and Reynolds and Marsico (1993), which combine in situ and satellite-derived SST data using Poisson's equation to produce "blended" products, with an analysis of the sea-ice edge as one boundary at −1.8°C. The in situ SST data used consist of quality controlled ship and buoy observations available over the Global Telecommunication System (GTS). Satellite data are obtained from the Advanced Very High Resolution Radiometer (AVHRR) on National Oceanic and Atmospheric Administration (NOAA) polar orbiting satellites. The

SST retrievals are produced operationally by NOAA's Environmental Satellite, Data and Information Service (NESDIS) and are available beginning in November 1981. The global coverage provided by satellite estimates of SST is a considerable advantage over the sparse coverage of in situ data, and satellites also provide useful information about patterns and gradients of SSTs. The absolute accuracy of satellite-derived SST, however, is uncertain; substantial corrections are necessary where in situ data are available to provide calibration (Reynolds 1988). Without real-time bias corrections, SST analyses using operational AVHRR retrievals are not useful for climate monitoring or climate modeling. A disadvantage of the blending technique to correct biases in the satellite data relative to the in situ data, however, is the considerable degradation of the spatial resolution of the SST analysis.

With the OI product, the high resolution of the satellite data is better preserved and the analysis is done weekly (and daily for operations). The first step is to use the blending technique to provide a preliminary large-scale time-dependent correction of satellite biases. The in situ and bias-corrected satellite SST data are then analyzed using OI on a 1° latitude and longitude grid. Optimal interpolation produces an interpolated value from a weighted sum of the data. Weights are computed using estimates of local spatial covariance and data error variance. The first guess is the previous analysis of the anomalies, which, therefore, persists the anomalies in the absence of new information. The technique does not otherwise utilize information from earlier or later times.

The NCEP OI product is global and is therefore very useful for both climate monitoring and as a lower boundary condition for AGCM simulations; however, because of its reliance on SST retrievals from the AVHRR instruments, its period of coverage extends only from November 1981 onward.

b. NCEP EOF analyses

To produce a near-global SST dataset based on in situ data farther back in time, Smith et al. (1996) developed an interpolation method that takes advantage of the full covariance structure in the more recent OI SST field. Using 12 yr (January 1982–December 1993) of the monthly OI SST anomalies, determined by subtracting the adjusted OI climatology of Reynolds and Smith (1995), EOF spatial basis functions are computed for six subregions of the globe (see Fig. 2 and Table 3 of Smith et al. 1996). The dominant regional EOF modes are then fit to detrended

2° monthly median SST anomaly statistics from the Comprehensive Ocean–Atmosphere Data Set (COADS; Woodruff et al. 1987) to determine the time dependence of each mode. Regionally complete fields of monthly SST anomalies on the 2° COADS grid are reconstructed from the spatial modes, the subregions are combined to produce a near-global product, and finally the smoothed long-term trend is restored at each grid point. The number of EOFs retained for each subregion varies from 16 to 25, and they generally explain between 80% and 90% of the variance in the OI. The maximum number of EOFs are selected in an attempt to minimize the data noise but maximize the reconstructed signal.

The EOF-based SST analyses have a southern limit of 45°S because of the relative lack of in situ SST observations at high southern latitudes, and their northern limit (~65°N) is determined by regions of sea ice that are not well represented by the EOFs. Grid points not reconstructed are assigned values from the 1982–93 OI climatology of Reynolds and Smith (1995). The period of coverage is from January 1950 to the present. Overall, Smith et al. (1996) concluded that their EOF-based interpolation method [or eigenvector projection method; see Kaplan et al. (1997), (1998)] results in an improved SST analysis that more realistically represents the large-scale SST structure in sparsely sampled regions than more traditional analysis techniques. In regions where in situ sampling is dense, the EOF-based reconstruction does not have such a clear advantage.

c. GISST analyses

The GISST SST analyses have complete global coverage and are designed explicitly for forcing climate models. Several different versions exist and updates are frequent. The version examined here is GISST 2.3b, which updates GISST 2.2 described by Rayner et al. (1996). Total SST fields are available month by month on a 1° grid, although intermediate processing of the anomalies is done on coarser grids.

Several different steps have been used to construct the GISST analyses.

1) *GISST 2.1*. The analyzed fields are monthly from January 1982 onward and make use of the Poisson blending technique of Reynolds (1988) to incorporate bias-corrected satellite-derived SST data from the AVHRR instruments with in situ data from the quality controlled U.K. Meteorological Office historical SST dataset [MOHSST version 6; Bottomley et al. (1990); Parker et al. (1994); Parker

et al. (1995a)] and COADS, sea-ice data, and a statistically based ice-zone SST specification (Rayner et al. 1996). The analysis is done using an anomaly resolution of 2°, and total SST fields are obtained by adding the 1° resolution 1961–90 climatology of Parker et al. (1995b).

2) *GISST 1.2*. Spatial gaps in 5° gridded monthly MOHSST (version 5) anomalies for 1903–48 are infilled using coarse-resolution (~10°) ocean basin EOFs based on seasonal data from 1901 to 1990. For data-poor areas where EOFs cannot be adequately defined, SST values are estimated using the Poisson equation technique with assumptions about SST near the ice edge as in Parker et al. (1995c). The anomalies used and the 1° resolution climatology added back are with respect to 1951–80.

3) *GISST 2.2*. Over the period 1949–81 SST fields are based on reconstructions using eigenvectors of in situ anomalies with a 2° spatial resolution in a fashion similar to that employed in the NCEP EOF analyses of Smith et al. (1996). Details differ, however, in the length of the SST records used to construct the EOFs, in the areas over which the EOFs are computed, and in the methods used to deal with the trend component. The first step of the analysis is to remove a global multidecadal SST trend signal represented by the first global EOF of low-pass filtered coarse-resolution data (Fig. 3 in Rayner et al. 1996; see also Parker and Folland 1991). Next, 2° resolution regional EOFs over four ocean basins are calculated using MOHSST (version 6) anomalies over 1951–90 interpolated using a background field reconstructed from 20 coarse-resolution global EOFs. As for the NCEP EOF analyses, the areas covered by the regional EOFs overlap slightly (Rayner et al. 1996), so a near-global SST reconstruction is made by averaging across the overlapping domains before adding back the low-frequency-trend EOF. Where data coverage is too sparse to determine EOFs (e.g., parts of the Southern Ocean and the southeastern Pacific), SSTs are infilled using Laplacian interpolation similar to the scheme used in GISST 1.2. As for GISST 2.1, ice-zone SSTs are specified statistically, and the 1° resolution 1961–90 climatology of Parker et al. (1995b) is added back to obtain the total SST fields.

The combination of the EOF-reconstructed SSTs over 1949–81, GISST 1.2, and GISST 2.1 is termed GISST 2.2. Rayner et al. (1996) discuss im-

provements relative to GISST 1.1 (Parker et al. 1995c): in particular, after 1948 GISST 2.2 contains more accurate sea-ice data, a better representation of near-ice SST, an improved background climatology, and higher-resolution SST analyses.

4) *GISST 2.3.* The SST analyses over 1903–48 are improved and extended back to 1871 by reanalyzing the data in a similar way to the EOF-reconstructed SSTs discussed above. In particular, MOHSST (version 6) data are used throughout, a global trend EOF and 4° resolution ocean basin EOFs are utilized in the reconstructions, and the anomaly analysis is done on a 4° grid.

At high latitudes, marginal ice-zone SSTs in GISST 2.3 are obtained through simple regression relations between SST and sea-ice concentration in areas where both exist. The sea-ice concentration data are from Walsh (1995) over the Arctic and various climatologies and other sources over the Antarctic and inland seas (see Rayner et al. 1996 for details). Observed monthly mean SSTs from in situ sources for 1961–90 over the Northern Hemisphere (NH) and satellite-derived SSTs for 1982–94 over the Antarctic are regressed against the sea-ice concentration data. The resulting equations depend on season and longitude over the Arctic, but only season over the Antarctic.

Satellite data used in the construction of the GISST 2.3 analysis were erroneously biased from 1982 onward in the initial release of the data (GISST 2.3a). The correction of this error led to the release of GISST 2.3b, which is the dataset examined here.

d. LDEO analyses

The global analyses of Kaplan et al. (1998) are derived from 5° in situ data from MOHSST (version 5; Parker et al. 1994) using a statistical method known as reduced space optimal smoothing. The period of record is 1856–1991, although an unpublished update based on COADS data through the end of 1997 was kindly provided to us by Y. Kushnir (1998, personal communication). Since the resolution of the data is lower than either of the NCEP products or GISST, the LDEO SST analyses are not as useful as a lower boundary condition for AGCM integrations. They have been used, however, for examining low-frequency modes of global SST variability (e.g., Enfield and Mestas-Nuñez 1999; Cane et al. 1997).

The OS method and its differences from other statistically based analysis methods such as OI and the eigenvector projection technique of Smith et al. (1996) are described in detail by Kaplan et al. (1997), who also demonstrate the differences in the analyses over the relatively data-rich Atlantic north of approximately 30°S. The techniques are then applied and compared over the global oceans by Kaplan et al. (1998).

The OS technique combines data reduction and least squares optimal estimation. The data reduction involves computing EOFs of the MOHSST data over the period 1951–91, then using a subset (80 global EOFs) as a basis for the analyzed solution. Unlike the approaches used in the construction of the NCEP EOF and the GISST analyses, the trend component is not analyzed separately in the OS technique, so unless the base period contains the whole trend, it is likely to be underestimated.

It is the determination of the stationary spatial covariance of the SST field that is perhaps the most novel feature of the OS analysis. Because of data gaps and observational error in the covariance field, Kaplan et al. (1997) smooth it in each spatial direction in such a way as to preserve the large-scale relations in the original covariance while eliminating the small-scale variations, which are presumed to be dominated by observational error. The variance of the original SST data removed by this procedure is then recovered by inflating the variance in the smoothed spatial covariance. The EOFs to be used as a basis set are then calculated and are also used for fitting a first-order linear autoregression model of time transitions. Thus, the OS technique provides a best estimate of SST based on available observations at all space points, but it also utilizes information from all times (preceding, during, and after the analyses time), in contrast to the OI and projection methods.

Over periods of relatively good data coverage, Kaplan et al. (1997, 1998) find that the OS, OI, and projection methods give comparable results; however, at times of especially poor coverage the use of information from other times appears to give the OS method an advantage. As in the Smith et al. (1996) SST analyses, the OS SST product is not global: in extremely data-sparse areas no attempt is made to estimate the spatial covariance field.

e. SST climatologies

The adjusted OI SST climatology of Reynolds and Smith (1995) has recently been updated to the World Meteorological Organization (WMO) suggested base period of 1961–90. This new climatology is described by Smith and Reynolds (1998). Briefly, a 1° resolution SST climatology is formed from the NCEP OI

product over the 1982–96 period, which includes three more years of OI analyses than were in the older climatology of Reynolds and Smith (1995). Next, the NCEP 2° EOF-reconstructed SST analyses are used to compute a monthly climatology over the desired WMO base period, with COADS data used to fill inland sea areas. This climatology is then used to adjust the higher-resolution OI climatology to the 1961–90 base period following the procedures in Reynolds and Smith (1995), so that equatorial upwelling and fronts remain well resolved. Absolute differences between the 1950–79 and 1961–90 adjusted OI climatologies are generally less than 0.2°C and appear to reflect real changes in the climate, such as colder SSTs over the North Pacific and northwest Atlantic associated with intensified Aleutian and Icelandic low pressure systems over the past 20 years (e.g., Hurrell 1996).

The other 1° resolution climatology we examine is the GISST 2.2 product described by Parker et al. (1995b), which improves upon earlier UKMO climatologies (e.g., Bottomley et al. 1990) especially in data-sparse regions because of the utilization of satellite data. A globally complete background SST field is first created from blended satellite and MOHSST data over 1982–94 (GISST 1.1; Parker et al. 1995c). Worldwide in situ SSTs over the years 1961–90, combined with statistically based estimates of SSTs in sea-ice zones, are then blended with the background SST field as outlined in Parker et al. (1995b), and the resulting monthly SSTs are averaged to form a new monthly 1° resolution climatology representative of the WMO standard reference period. This monthly climatology is then interpolated to daily resolution, and these values are used in the quality control of the latest version of MOHSST (version 6), which, in turn, is used in the development of the monthly GISST 2.2 analyses described above. The final step is to average the resulting GISST 2.2 analyses to form a new GISST 2.2 1961–90 monthly climatology.

3. AGCM experiments

General circulation models of the atmosphere forced over time with observed SSTs are an important tool in our ability to understand climate variability and predictability. Knowledge of the observed SST field is also critical for the analysis and reanalysis of atmospheric data. Relatively little attention has been paid, however, to the impact of different SST analyses on model simulations. Usually different models are forced with the same SSTs (e.g., as in AMIP), or one SST analysis is used to force the same model but with changes to other climate forcings. When a new SST analysis comes along, often the atmospheric model has changed over time as well so the impact of the different SSTs on the simulated atmosphere cannot be determined. Moreover, especially for the climate record of the past several decades, when SSTs are much better observed, the impact of differences in SST analyses is generally assumed to be small compared to the noise levels of internal atmospheric variability. One might be tempted, for instance, to increase ensemble size by averaging together integrations performed with the same model but forced with different SST analyses of the same period of time.

Here we examine the impact of differences in analyzed SSTs on a model-simulated climate with a recent version of the NCAR AGCM, CCM3, described in detail by Kiehl et al. (1998). The standard model configuration uses a triangular wavenumber 42 (T42) horizontal spectral resolution (approximately a 2.8° by 2.8° transform grid) with 18 unequally spaced vertical (hybrid) levels. Fifteen integrations are analyzed. One five-member ensemble is forced with the monthly NCEP EOF-reconstructed SST analyses over 1950–97, while another five-member ensemble is forced with those SSTs through 1981 and the NCEP OI SST analyses over 1982–97. The final five runs are forced with the GISST 2.2 SST analyses over the period 1903–94. Analyzed monthly SSTs are assigned to the midmonth date and updated every time step at each ocean grid point using linear interpolation.

4. SST comparison results

a. Climatological SSTs

We compare the NCEP-adjusted OI climatology of Smith and Reynolds (1998) to the UKMO GISST 2.2 climatology of Parker et al. (1995b), which are both representative of the 30-yr period 1961–90. Month-to-month differences between the two climatologies reveal many of the same features; therefore, a summary is given by the annual mean of the monthly differences (Fig. 1). Overall there is reasonably good agreement, with absolute differences less than 0.25°C over most of the global oceans. The largest differences (~2°C) are at high latitudes where the GISST climatology is warmer. These differences stem from the extremely low number of in situ observations in these regions and the very different methodologies employed at NCEP

and UKMO to estimate SST near sea ice, as described earlier. In the high southern latitudes and in polar regions, where sampling is extremely poor, both analyses are questionable.

Other large differences relate to the ability of the NCEP-adjusted OI climatology to resolve real small-scale structures and sharp SST gradients. This is particularly evident in the equatorial Pacific upwelling region, where the GISST climatology is warmer by almost 0.5°C, and in areas where SST gradients are large. Absolute SST differences exceed 1°C, for instance, in the retroflection region south of Africa and near the Peru, Falkland, and Benguela Currents. Even more striking are the large differences in the Kuroshio extension in the North Pacific and in the Gulf Stream of the North Atlantic. In the latter area, a narrow Gulf Stream is well defined in the NCEP climatology, but not in the GISST, which leads to a dipole structure in the differences with absolute values exceeding 1°C (Fig. 1).

FIG. 1. Differences between the 30-yr 1961–90 climatologies of GISST and NCEP (GISST–NCEP) in °C. The color contours are every 0.25°C up to absolute 1°C, then 0.5°C up to absolute 3°C.

b. Local reproducibility of SSTs

We compare the four monthly SST datasets described in section 2 using simple standard statistics. Correlation coefficients, root-mean-square (rms) differences, standard deviations, linear trends, and lag-1 month autocorrelations are computed after removing separate annual cycles using the monthly means, thereby eliminating systematic biases.

Higher-resolution SST datasets, such as the NCEP OI and GISST 2.3b products, are directly compared; however, they are area averaged onto coarser-resolution grids for comparisons with the NCEP EOF-reconstructed and LDEO SST datasets. For brevity, we first focus primarily on the NCEP OI and GISST 2.3b SST analyses from 1982 onward. Both incorporate in situ and satellite SST data over this period and, because they have complete global coverage, they are commonly used to force AGCMs.

The standard deviation of monthly SST anomalies from both the NCEP OI and GISST analyses is largest along the equatorial tropical Pacific a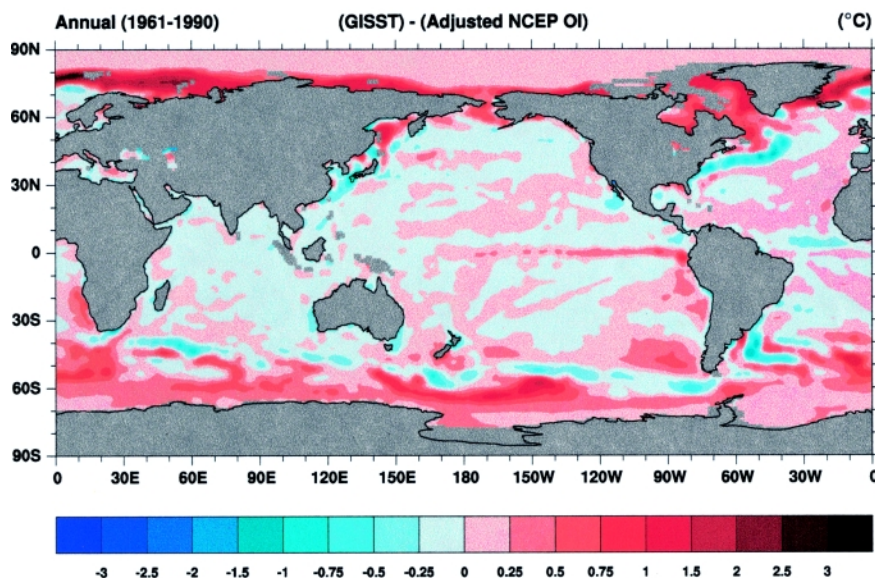nd South American coast where interannual variability associated with the El Niño–Southern Oscillation (ENSO) phenomenon is most pronounced (Fig. 2). Large month-to-month variability is also evident in the SSTs over the North Pacific and over the North Atlantic associated with the Kuroshio extension and the position of the Gulf Stream. In all of these areas the variance is larger in the NCEP OI analyses than in the GISST product partially reflecting the coarser resolution of the latter. Over much of the rest of the global oceans, however, monthly SST variability is greater in GISST. This is especially true over data-sparse regions where GISST relies heavily on locally interpolated in situ observations and sampling uncertainty is large. Smith et al. (1996) have also compared these two analyses to their EOF-reconstructed SSTs and our findings (figures not shown) are consistent. In general, the NCEP reanalyzed SSTs retain most of the variance of the OI product, with very good agreement over the northern oceans and slightly less variance over the tropical Pacific.

Global maps of correlation coefficients and rms differences between the NCEP OI and GISST 1° global SST monthly anomalies (Fig. 3) reveal that over the northern oceans and the eastern tropical Pacific correlation coefficients are highest, mostly exceeding 0.9. Values are generally lower than 0.75 south of about 10°N and are much lower locally over the western tropical Pacific and the southern oceans, both regions where the number of in situ observations drops off considerably. Root-mean-square differences be-
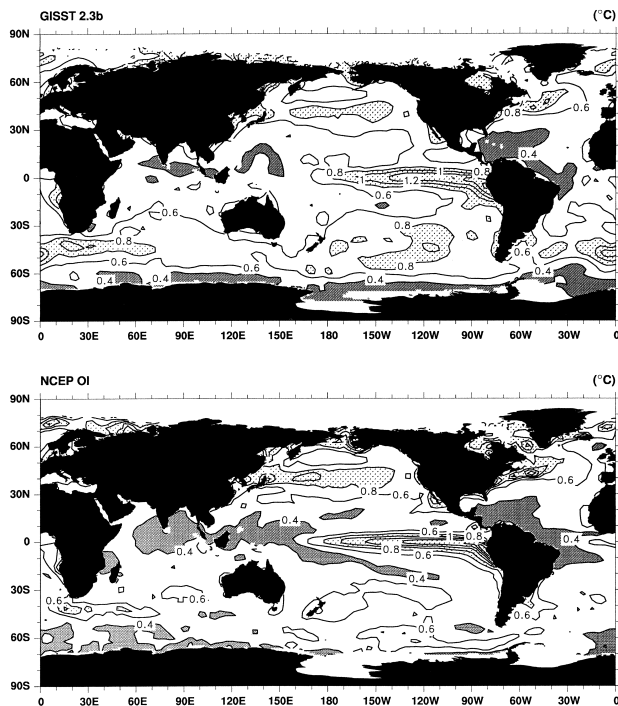
FIG. 2. Standard deviations of monthly SST anomalies for 1982–97 with a contour interval of 0.2°C for GISST 2.3b and NCEP OI. Values greater than 0.8°C are stippled and values less than 0.4°C are shaded.
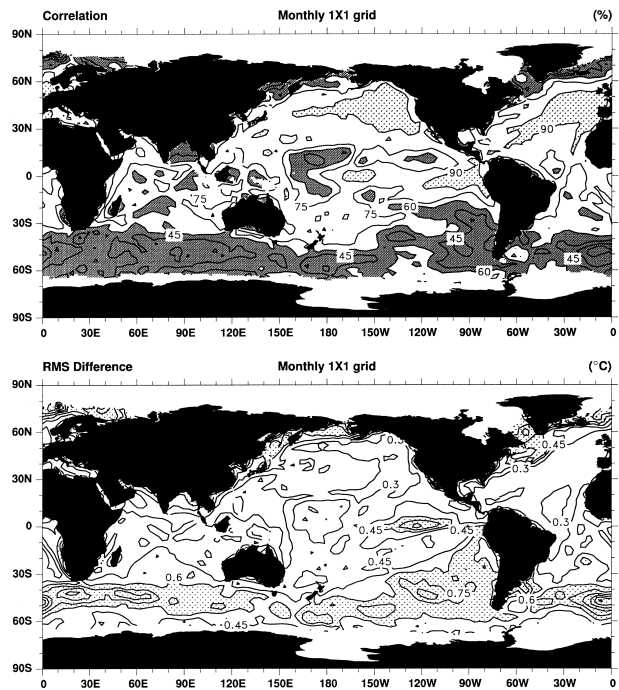

FIG. 3. Correlations (top) and rms differences (bottom) between monthly SST anomalies from GISST 2.3b and NCEP OI for 1982–97. The correlations are contoured every 15% and values below (above) 0.6 (0.9) are shaded (stippled). The rms differences are contoured every 0.15°C and values exceeding 0.6°C are shaded.

tween the SST analyses increase from less than 0.2°C over the central North Atlantic to over 0.6°C over the eastern tropical Pacific, in the eastern Pacific south of 10°S, and generally south of 35°S except near New Zealand (Fig. 3b). In the two latter areas the correlations are less than ~0.6. As shown in TCH, there is a striking resemblance between both the pattern of correlations and rms differences and the numbers of in situ observations available. Because essentially the same in situ and satellite observations are used in both the GISST and NCEP OI products, the correlations and rms differences reflect differences in the quality control and analysis methods, and it is apparent that there is uncertainty in the true anomalies related to the rms differences (see discussion in section 6).

One reason for the differences is the coarser resolution used in the GISST analysis. A coarser-grid analysis may be more realistic for the presatellite record, and especially before 1950 when sampling is often poor. But it is costly when high-resolution data are available. For example, GISST cannot resolve equatorial upwelling as well as the NCEP OI (Fig. 1). In addition, the GISST analysis relies more heavily on in situ data, which are noisier than satellite data (e.g., Reynolds and Smith 1994).

Another reason for the differences in correlation relates to the size and persistence of the climate signal, and some insight into this is given by maps of the standard deviation of the monthly anomalies (Fig. 2). This quantity squared depicts the sum of the actual signal plus the noise variance. Although details differ, large signals are hard to miss and are captured in both analyses. This is true, for instance, over the eastern tropical Pacific where the El Niño signal is large. In contrast, over much of the rest of the tropical oceans, the signal is small and the influence of noise is greater.

c. Persistence of anomalies

Significant differences between the NCEP OI and GISST analyses are also evident in the persistence of SST anomalies from one month to the next (Fig. 4). Relatively large (> 0.7) lag 1-month autocorrelations are evident in the NCEP OI SSTs over much of tropical and North Pacific Oceans, over the tropical and subtropical Atlantic, and generally south of about 50°S. Somewhat lower values (< 0.6) are associated, for instance, with the major ocean currents off the coasts of continents such as the Kuroshio and Gulf Stream where cold and warm ocean rings form and eddy activity is known to be large. These values seem
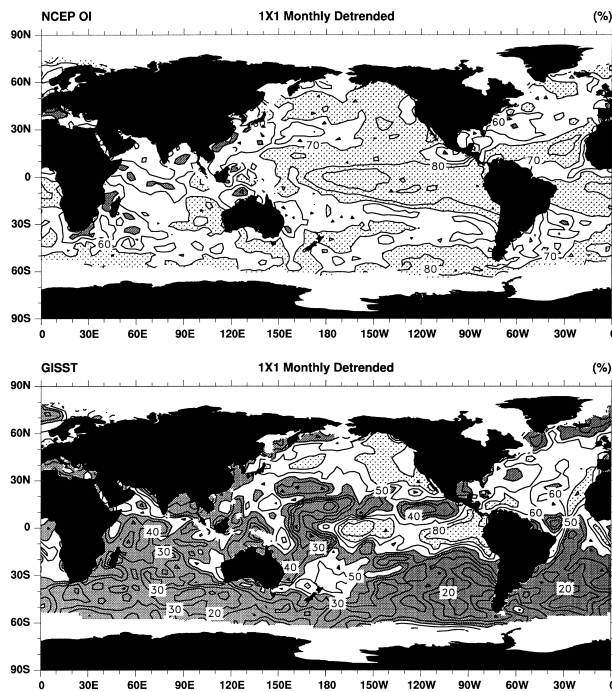
FIG. 4. Lag-1 month autocorrelations of SST anomalies for the period 1982–97 for the NCEP OI (top) and GISST 2.3b (bottom). Both series were detrended before computation, although this makes little difference. Values greater than 70% are stippled and less than 50% are shaded, and the contour interval is 10%.
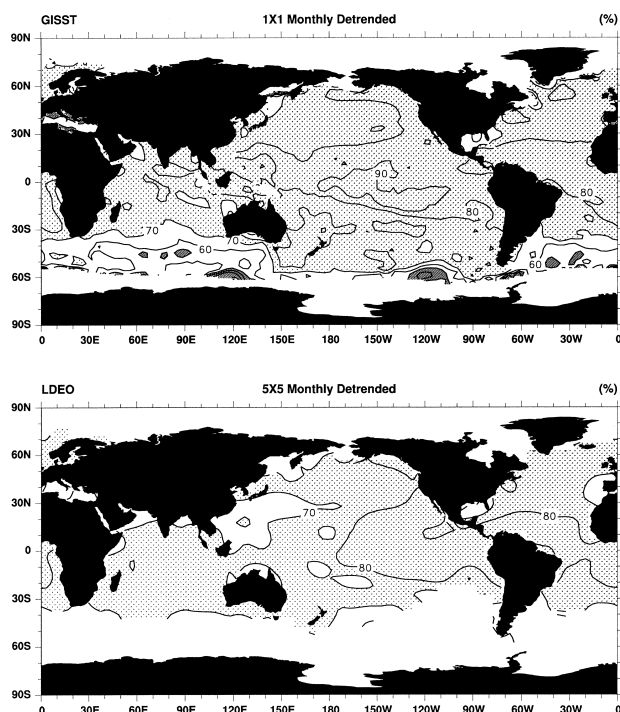
FIG. 5. Lag-1 month autocorrelations of SST anomalies for the period 1950–81 for GISST 2.3b (top) and LDEO (bottom). Both series were detrended before computation. Values greater than 70% are stippled and less than 50% are shaded, and the contour interval is 10%.

to be very reasonable and are not a function of data density. In sharp contrast are the lag-1 month autocorrelations in the GISST 2.3b analyses, which are much lower nearly everywhere. Values less than 0.3 are widespread south of about 30°N, except over the tropical Pacific and portions of the northern subtropical Atlantic. Over the latter region a band of relatively high correlations in the GISST analyses extends from North Africa to Brazil along one of the world's major shipping routes. For the GISST autocorrelations there is a striking relation with the number of in situ observations (e.g., see Fig. 8 in TCH), with values falling off in regions with little data.

The low lag-1 month autocorrelations and their pattern in GISST 2.3b from 1982 onward relates primarily to the use of the Poisson blending technique (N. A. Rayner 1998, personal communication). From 1982 to 1994 in situ data were infilled using the Laplacian of the AVHRR SST field (after 1995 the NCEP OI SSTs were used for this purpose), and the result is that areas of poor in situ coverage are temporally incoherent. As expected, this problem can be partially alleviated through smoothing. When the lag-1 autocorrelations are computed from running 3-month

mean anomalies, for instance, the agreement between the NCEP OI and GISST products improves considerably, with autocorrelations exceeding 0.7 at all ocean points in both datasets, although autocorrelations remain slightly lower in the GISST product (not shown).

The lack of temporal continuity in the monthly GISST analyses after 1981 is not, however, evident when earlier times are examined. Over the period when the GISST fields are based mainly on reconstructions using eigenvectors of in situ SST anomalies, for example, lag-1 month autocorrelations are much higher (Fig. 5). In fact they seem to be too high in the Kuroshio extension and Gulf Stream regions (cf. the NCEP OI in Fig. 4). Low values are still found locally over mid- and high latitudes of the Southern Hemisphere (SH), where there is no reason physically for the values to be lower than comparable NH regions where a Sverdrup balance dominates. This again reveals that the data coverage is a factor in the analyses.

Also shown in Fig. 5 is the autocorrelation from the LDEO analyses. For direct comparison with the GISST results, the lag-1 month autocorrelations are shown for the period 1950–81. The results are consis-

tent with those from other periods as well and are very smooth, which reflects the coarse 5° resolution. Also the coarse resolution reduces the variability of small-scale eddies, such as in the Gulf Stream, and thus increases the autocorrelation. Otherwise the values are reasonably consistent with those of the NCEP OI analyses.

d. Multidecadal trends

Trends in the analyses over the post-1950 period are quite similar, as would be expected from the common database and the fact that this period or part of it is used to define the statistics for the EOF infilling or optimal smoothing. In contrast, for longer periods, the trends are quite different. Figure 6 presents the linear trends for 1900 to 1997 from the GISST and LDEO analyses, both of which are based upon mostly the same data (and in particular with the same corrections for observing methods). The differences are substantial and trends are generally more negative in the LDEO dataset, including cooling trends in the tropical eastern Pacific, as reported by Cane et al. (1997) (here the data have been updated), in the subtropical North Pacific and South Pacific, and more extensively in the North Atlantic.
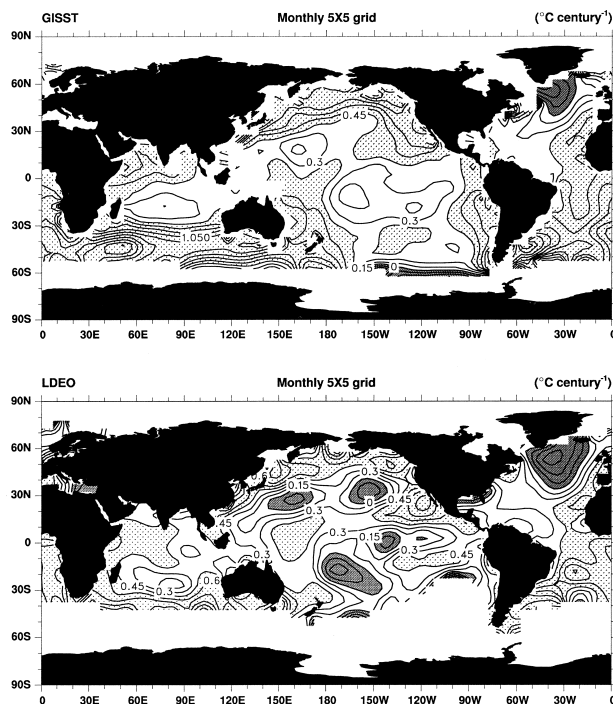

FIG. 6. Linear trends of monthly SST anomalies for 1900–97 on a 5° grid for GISST 2.3b and LDEO. The contour interval is 0.15°C per century; values greater than 0.45 are stippled and negative values are shaded.

While the GISST results are not guaranteed to be correct, the trend is analyzed separately, as it is desirable to take into account the nonstationary effects trends produce in any analysis of variance. The low-frequency signal is also removed in the NCEP analyses of Smith et al. (1996, 1998) before gridding of the residual signal, in order to avoid this problem, and the low-frequency variance is only added back on after gridding. In contrast, it is clear that the trend will not be correctly projected using the OS technique, unless the base period of 1951–91 contains the entire trend, which it clearly does not.

e. Area-averaged time series

It is often argued for climate purposes that temperature anomalies are large in scale so that averaging over larger areas better serves to define the anomalies while reducing sampling error. In the following, area averages over the extratropical NH, extratropical SH, and the Tropics are taken to emphasize the regional variations and to see the extent to which the different SST analyses agree. The latitudinal bounds are 45°S and 60°N, which correspond to the limits of the NCEP EOF-based SST analysis. The comparisons are made with monthly anomalies relative to 1950–79.

Over the NH extratropics (20°–60°N; Fig. 7) monthly SST fluctuations are highly correlated between the GISST and both NCEP analyses (0.87) over the period since 1950, and the largest differences tend to appear in individual months rather than over extended periods. Monthly differences become smaller after the mid-1970s except toward the end of the record (e.g., May 1995 and August 1996) because of spiky behavior in the GISST analyses. Of note are the differences in warming over the past five years, with NCEP EOF SSTs showing distinctly less warming.

To examine the extent of the agreement and discrepancies in more detail, 5-yr running means of cross correlations, autocorrelations, standard deviations, and rms differences were computed. One noticeable feature this revealed over the NH was very low correlations (~0.2) between the GISST and NCEP EOF SSTs during the early 1960s and during the mid-1970s, and these are primarily due to very different trends over those short periods. Centered on January 1961, for instance, the GISST analyses cool at a rate of −0.16°C per 5 yr while the linear trend in the NCEP data is 0.11°C per 5 yr (Fig. 7). The NCEP EOF SSTs exhibit slightly less monthly variance than either the GISST or the OI products, consistent with the findings of Smith et al. (1996), although this characteristic is more

noticeable in regions where the in situ coverage is worse, such as the Tropics and the SH.

The best agreement in all statistical measures among the three datasets occurs when the SST anomalies are averaged over the Tropics (Fig. 7). Over this portion of the globe (20°S–20°N), the large interannual variability associated with ENSO is well captured in each analysis, although absolute differences of up to 0.2°C are evident between the GISST and NCEP reconstructed data prior to the 1980s. Clearly seen in all three products is the post-1976 jump to higher SSTs (Trenberth and Hoar 1996, 1997). Over the past 15 years, differences are small but show a slight warming in the GISST data relative to both NCEP products.

The agreement between monthly SST anomalies is worst over the SH extratropics (20°–45°S) where absolute differences are large (up to 0.5°C between the GISST and NCEP OI analyses, which correlate at 0.52 since 1982) and spiky in character (Fig. 7). The warming of the GISST relative to both NCEP products is most pronounced in the SH, which is also where there is the least amount of in situ data. As discussed in section 6, it is likely that the NCEP OI product is biased cold, and even more so in the 1990s, because of the satellite data. The NCEP EOF analyses do not use satellite data, however, so the differences with GISST must be due to the processing of in situ data (R. W. Reynolds 1999, personal communication).

5. Tropical Pacific SST differences and atmospheric impact

Within the Tropics there is a fairly direct tropospheric response to SST anomalies and masking by internal atmospheric variability is relatively small compared with the extratropics (Shukla 1998). It is over this portion of the globe, therefore, that the SST differences are most likely to produce a change in the atmospheric circulation above the noise of chaotic natural variability in AGCM simulations. A review of the tropical SST teleconnections (Trenberth et al. 1998) indicates that the signal is more likely to be found if the SST anomalies last for a season or longer.

The temporal evolution of SST anomalies over the large 1982–1983 El Niño event is depicted differently in the NCEP OI analyses relative to the GISST analyses (Fig. 8). Monthly tropical Pacific SST anomalies, referenced to a 1950–79 base period, are shown as a function of time and longitude averaged between 5°S and 5°N. The lack of temporal consistency in the
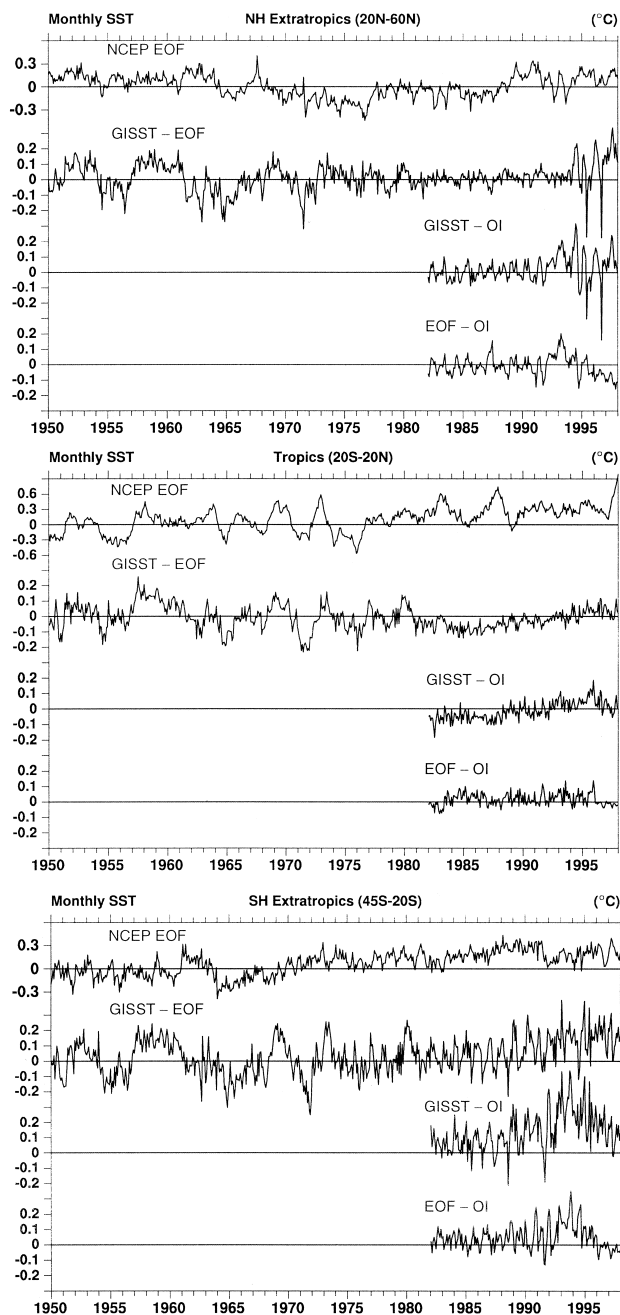

FIG. 7. Monthly time series for large area averages for the NH extratropics (20°–60°N), Tropics (20°N–20°S), and SH extratropics (20°–45°S) in °C. Shown are the SST time series from NCEP EOF and the differences (GISST–NCEP EOF), (GISST–NCEP OI), and (NCEP OI–NCEP EOF).

monthly GISST data is clearly evident: relatively large SST anomalies of one sign are often followed 1 or 2 months later by equally large SST anomalies of opposite sign, in contrast to the relatively smooth temporal behavior of the NCEP OI product. Differences between the two total SST fields (Fig. 9) further re-
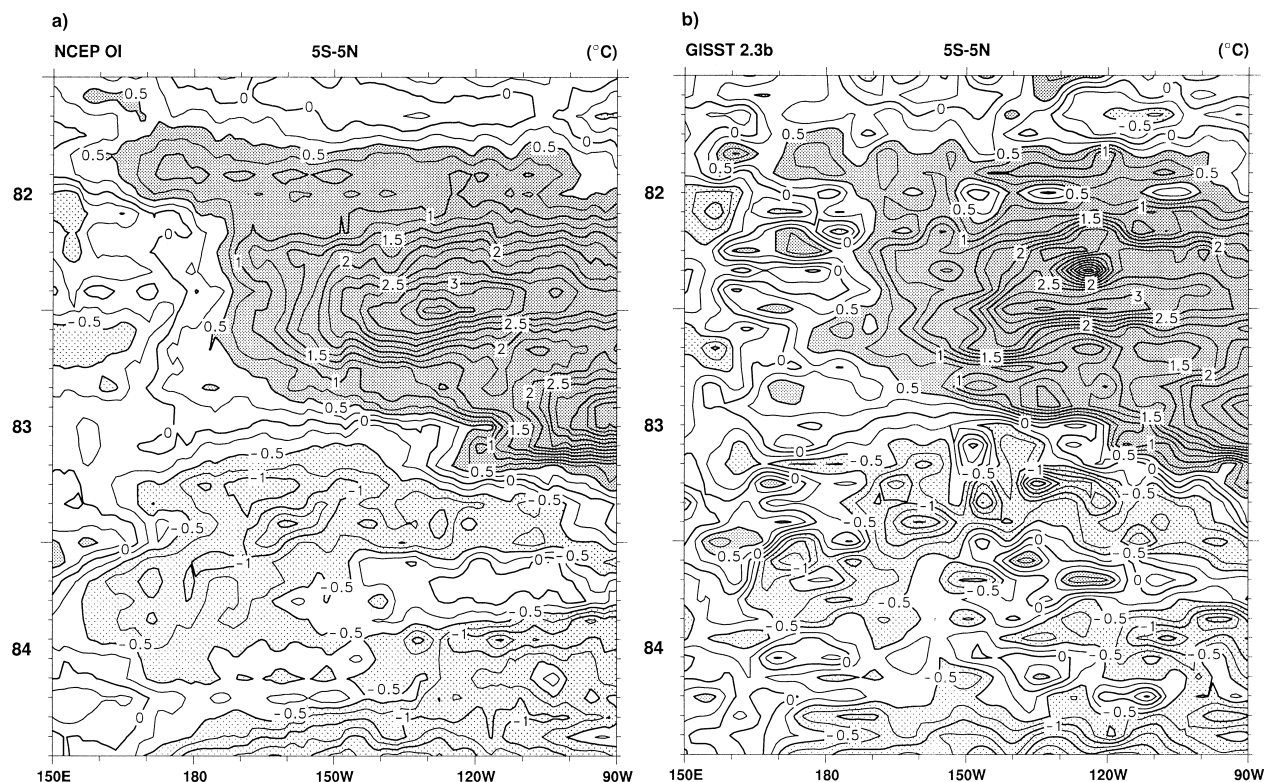
FIG. 8. Longitude–time sections for the period from Jan 1982 through Dec 1984 of the tropical Pacific averaged from 5°N to 5°S for (a) NCEP OI and (b) GISST 2.3b. The contour interval is 0.25°C and values greater than 0.5°C are shaded and less than −0.5°C are stippled.

veal the noisier structure in GISST. Some of this can be removed by smoothing (e.g., 3- or 5-month running averages), but systematic differences remain.

We have examined the impact of differences between the NCEP OI, the NCEP EOF reconstructed, and the GISST 2.2 SST datasets on five-member ensemble simulations performed with CCM3. Differences between the ensemble-mean monthly total precipitation averaged over 5°S–5°N (Fig. 9) are as large as 5 mm day^{-1}, especially west of the date line, and the noisier structure from the GISST analysis is apparent. The relationship between differences in total precipitation and SST is, however, nonlinear. Over the eastern tropical Pacific, for instance, differences between two SST analyses as large as those in Fig. 9 do not usually translate into significant precipitation differences except during warm events. This is seen by the lack of rainfall differences after mid-1983 when the SSTs returned to below normal in this region (Fig. 9). Thus over cold upwelling equatorial waters it does not rain in the model (or in nature) even if one SST analysis is 1°–2°C warmer or colder than another. The same is not true over the warm pool region, however, where ensemble-mean monthly total precipita-

tion differences over 1982–84 (and all other periods as well) are more frequently (but not always) of the same sign as the SST differences, and this is also true when the El Niño warming spreads into the eastern Pacific (as in 1982–early 1983) (Fig. 9).

How realistic is the precipitation response in CCM3 to differences in analyzed SSTs? The correlation coefficients from 1979 to 1995 between monthly SST anomalies from Smith et al. (1996) and monthly precipitation anomalies from Xie and Arkin (1996) reveal that the highest values (> 0.45) are observed over the tropical Pacific, with considerably lower values elsewhere but with large-scale structure (Fig. 10). In comparison, the gridpoint correlations between the same SST dataset and the ensemble-mean precipitation anomalies from CCM3 reveal higher correlations that are positive almost everywhere (Fig. 10). Similar results have been found for other AGCMs as well, including those used in the reanalysis of atmospheric data (Masutani 1997). This most likely indicates a shortcoming of most state-of-the-art AGCM experiments driven with specified SSTs. In such integrations heat fluxes from the ocean into the atmosphere do not cool the SSTs (Saravanan 1998), whereas, over the

extratropics, it is established that the atmosphere typically drives the ocean rather than the other way around (Trenberth et al. 1998). Because there are substantial uncertainties in the precipitation dataset (Xie and Arkin 1996), correlations may be lower in the "observed" panel in Fig. 10 than in reality. Nevertheless, it is likely that these kinds of AMIP runs are only pertinent for SST anomalies in the Tropics where the ocean drives the atmosphere and has a large upper ocean heat content support. A salient point here is that most AGCMs, including CCM3, do respond relatively strongly and realistically to changes (or differences) in monthly SST in low latitudes.

We found many examples of this using CCM3, and we illustrate one in Fig. 11. Averaged over the three months December–January–February (DJF) 1954/55, SSTs in GISST near and just west of the date line along the equator are 1°C warmer than they are in the NCEP EOF-reconstructed SST analyses. Differences in CCM3 en-
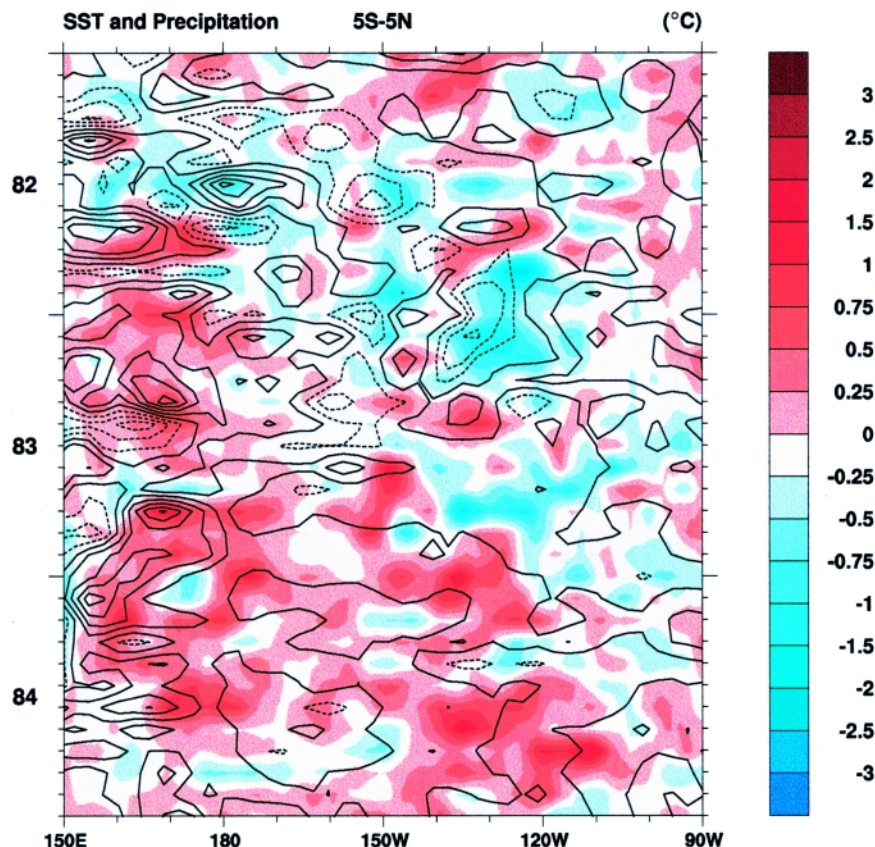

Fig. 9. Longitude–time sections for the period from Jan 1982 through Dec 1984 of the tropical Pacific averaged from 5°N to 5°S. In color is shown the differences between the SST anomalies (GISST–NCEP OI, from Fig. 8) with the color scale (right) in °C. The contours are the differences in ensemble-mean precipitation from CCM3 (GISST–NCEP OI runs) with a contour increment of 1 mm day^{-1}. Dashed contours are negative.

semble-mean total precipitation over this area reach up to 8 mm day^{-1} (with the GISST-forced precipitation rate greater), and such large changes in the modeled latent heating also imply changes in atmospheric responses and teleconnections.

Also shown in Fig. 11 are the differences between the ensemble-mean 200-mb divergent wind component and the 200-mb streamfunction averaged over DJF 1954/55. The response in the local divergent wind is clear and is as expected: over the region where the GISST SSTs are warmer and the CCM3 simulated precipitation rate is greater, the divergent outflow is stronger in the GISST-forced runs, and this is true in each individual member of the ensemble as well. Moreover, differences in the ensemble-mean subtropical and extratropical rotational flow are consistent with the changes in the tropical heating and outflow (Trenberth et al. 1998). In addition to twin anticyclonic centers just poleward of the region of the largest dif-

ferences in SST, strong wave trains are evident over the extratropical Pacific and downstream over both hemispheres with geopotential height anomalies exceeding 100 m in places. The magnitudes of the streamfunction anomalies are similar to those observed during warm or cold ENSO events (e.g., see Fig. 39 in Hurrell et al. 1998).

More generally, however, we found it difficult to identify an unequivocal extratropical response to analyzed SST differences using CCM3, as would be expected given the large amount of internal atmospheric variability in mid- and high latitudes. This is especially true of periods when large differences in tropical SSTs persisted for only 1 or 2 months, although such cases undoubtedly add spurious variance to the simulated climate. In several cases when tropical SST differences persist for several months or longer (such as DJF 1954/55), however, extratropical responses very consistent with the tropical heating anomalies that occur in nature are identifiable. This has important implications
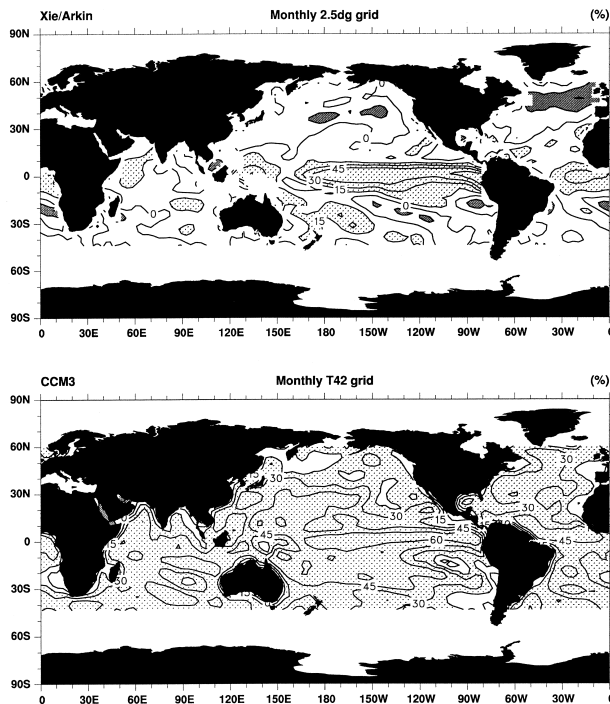
FIG. 10. Correlation coefficient between the local values of monthly mean precipitation and NCEP EOF SST anomalies. (top) Values for observational estimates from Xie–Arkin and (bottom) ensemble-mean precipitation from CCM3. The contour increment is 15% and values greater than 15% are shaded.

for the interpretation of climate variability and predictability from AGCM runs forced with analyzed SSTs.

6. Discussion and likely origin of discrepancies

We have shown that there remain significant differences between global SST datasets, both in their long-term climatologies and in the monthly anomalies that are important meteorologically. The results are similar to an earlier comparison of TCH and Folland et al. (1993), indicating that considerable uncertainty in our knowledge of the SST fields remains, in spite of improved analysis techniques and the greatly improved spatial coverage that satellite data provide. While much work remains to be done, progress is being made, the sources of many of the differences are known, and it is also reasonably well established how to fix many of the problems. No dataset is perfect; all have identifiable problems that can be addressed.

Using the COADS, TCH analyzed sources of errors for in situ SSTs. By assessing the variability within 2° latitude by 2° longitude boxes within each

month for 1979, TCH found that individual SST measurements are representative of the monthly mean to within a standard error of 1.0°C in the Tropics and 1.2–1.4°C in the extratropics. The standard error is larger in the North Pacific than in the North Atlantic and it is much larger in regions of strong SST gradient, such as within the vicinity of the Gulf Stream. This is because both within-month temporal variability and the within 2°–box spatial variability are enhanced. The total standard error of the monthly mean in each box is reduced approximately by the square root of the number of observations available. The overall noise in SSTs ranges from less than 0.1°C over the North Atlantic to over 0.5°C over the oceans south of about 35°S.

An additional problem with SST is that it is not as well defined as is desirable. Historically, SST has referred to a bulk near-surface ocean temperature measured by tossing a bucket over the side of a ship in order to obtain a water sample. The design and insulation of the buckets has changed with time, however, so that corrections must be applied (Folland and Parker 1995). During World War II, moreover, there was a switch from bucket measurements to measuring the temperature of water taken on to cool the ship's engines. These temperatures depend on the depth (3–7 m or more) and size (10–51 cm in diameter) of the ship's intake, the lading of the ship, the configuration of the engine room, and the point where the measurement is taken. Such differences are responsible for some of the noise in the SST measurements, but biases also arise because heat from the engine room more than offsets any cold bias from the depth of the intake. Overall, the difference between engine intake and bucket temperatures is typically 0.3°C (see TCH for a more complete review).

With satellite remote sensing of SSTs has come additional problems related to the skin (radiometric) temperature and the differences between the near-surface and bulk temperatures. While infrared satellite measurements of SST in principle give the skin temperature, many algorithms convert the skin temperature into a bulk SST measurement using a form of regression with selected buoy observations (e.g., Reynolds and Marsico 1993; Reynolds and Smith 1994). In the tropical western Pacific warm pool, Webster et al. (1996) compared the values of skin temperature versus the bulk temperatures at 1-cm, 0.5-m, and 5-m depth, where the latter corresponds to the typical depth of measurements from buoys and ship intakes. Skin temperatures are lower than the bulk 1-cm

depth SST by typically 0.2°C. In well-mixed windy conditions the three bulk SSTs are about the same, while in light wind conditions with high surface insolation strong near-surface warming occurs, and 1-cm temperatures are warmer than the 5-m depth SSTs by as much as 3°C. This gives rise to a significant diurnal cycle in SSTs on days with light winds. Webster et al. (1996) point out that a 1°C error in SSTs typically results in errors of 27 W m^{-2} in surface energy balance. This would be expected to have a significant impact on convection, as we have shown in the previous section.

As shown in section 4, in all regions there is a recent warming in GISST relative to the NCEP OI data and GISST is systematically warmer relative to the other analyses after 1982, especially in the Tropics and SH. These features are most likely related to differences in the processing of in situ data, and the cold bias in the NCEP OI product arises also from incompletely bias-corrected satellite data and is worse in the 1990s than in the 1980s (R. W. Reynolds 1998, personal communication). Reynolds (1988) showed that the satellite data were causing biases in the analyses, and the biases were different for daytime and nighttime retrievals. Folland et al. (1993) showed that the CPC satellite SST data were biased cold, typically by 0.5°C in the Tropics and SH, where in situ data used for bias correction are few. Satellite-based SSTs are not available in cloudy regions and, while they allow for water vapor attenuation, they are adversely affected by aerosols. The 1991 Mount Pinatubo eruption, for instance, produced a cold bias in the retrieved SSTs (Reynolds and Smith 1994). During the 1980s, in situ data from COADS were used to correct biases in the satellite-derived SSTs, so that roughly 30 000 ship observations per week globally were used. During the 1990s, however, only GTS in situ data have

been used, which translates into approximately 15 000 observations per week (R. W. Reynolds 1998, personal communication). Because the bias correction is underestimated where in situ observations are sparse, the difference in the number of ship observations is a likely cause of the cold bias in the NCEP OI analyses relative to GISST, which is especially evident over the SH mid- and high latitudes during the 1990s. Therefore, although global SST analyses are likely to be more complete when satellite-retrieved SSTs are included, they may be subjected to biases that vary in time.

Clearly there is scope for a more physical treatment of satellite measurements. It is desirable, for instance, to explicitly recognize the different nature of skin and bulk SST measurements and parameterize the diurnal cycle in satellite SST estimates, especially in light wind areas (Webster et al. 1996).
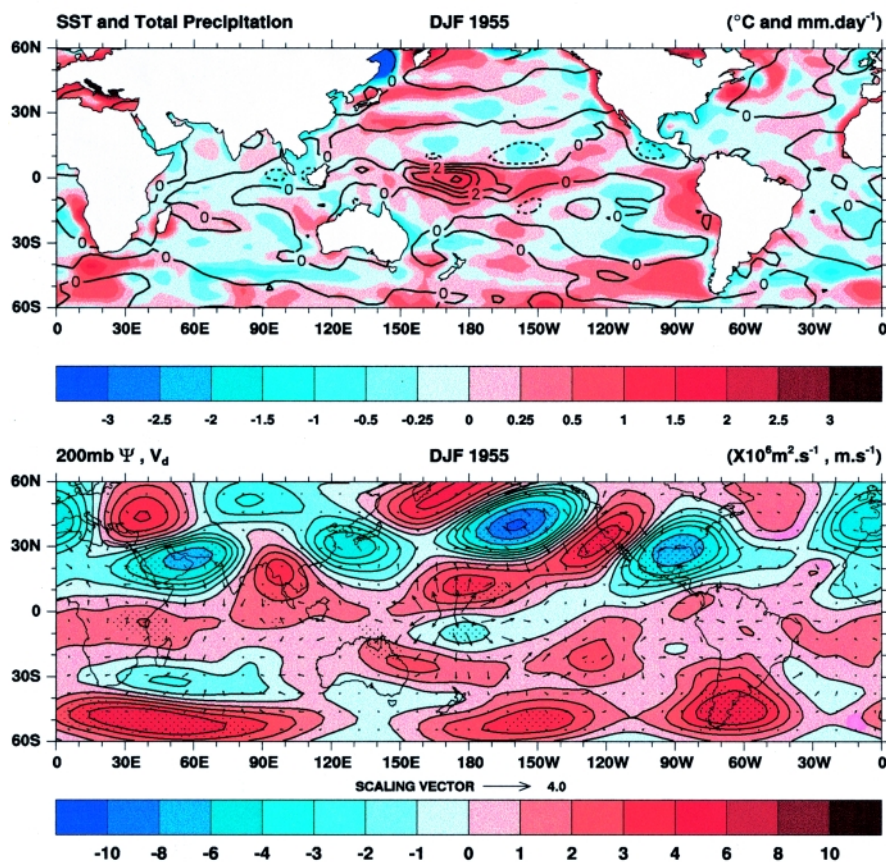

Fig. 11. For the season DJF 1954/55, (top) shown in color are the SST differences (GISST–NCEP EOF) in °C, and contoured are the differences in CCM3 ensemble-mean precipitation in mm day^{-1} with negative values dashed. (bottom) Differences in CCM3 ensemble-mean 200-mb divergent wind and the 200-mb streamfunction in 10^6 m^2 s^{-1}. Streamfunction values statistically significantly different from zero at the 5% level are stippled; negative values are in blue and positive values are in red. Note that the color scale is not linear. The longest divergent wind vectors slightly exceed 2 m s^{-1}.

Near the sea-ice zones large differences exist in the SST analyses that need to be resolved. Observations are few and far between in these regions but there is information that should allow improved empirical models to be developed and used to make the analyses of SST more reliable. In particular, it is important that the SSTs should be consistent with the observed ice cover. One of the most important problems is uncertainty in the sea-ice concentrations. During northern summer, for instance, climatological sea-ice concentrations over the Arctic based on objective analyses of microwave satellite observations differ by 20% or more in some regions from a subjective analysis (Knight 1984) of in situ and satellite data (R. W. Reynolds 1999, personal communication). The collocated data used to define the constants in the GISST regressions (section 2c and Rayner et al. 1996) primarily occur at low concentrations near the ice margins. The statistically generated SSTs at higher latitudes, however, are far from most of these data and are thus sensitive to the choice of the sea-ice concentration dataset. Careful comparisons and evaluations of different sea-ice analyses are therefore critical.

The problem of sparse or nonexistent data in the more distant past will remain. Quality control is also more difficult with few observations, and redundancy is essential for cross-checking. Our experience indicates that at least three observations are desirable before a monthly anomaly can be reliably defined. However, exploiting the full information contained in each observation and their time sequence is a problem that is only beginning to be addressed. The OS technique of Kaplan et al. (1998) takes full advantage of the spatial and temporal structure expected in the SSTs. The noise evident in the GISST analyses appears to come partly from shortcomings in quality control, which should otherwise catch the problems seen in Fig. 8. However, the lack of continuity in the GISST analyses after 1982 (Fig. 4) also shows the need to exploit the temporal persistence better.

As noted in the introduction and in section 4d, the treatment of trends is difficult and different approaches yield greatly different results. There is a need to recognize that a warming trend is present, for whatever reason, in the SSTs in most places. Moreover, a warming trend is expected from human activities and the increases in carbon dioxide and other greenhouse gases in the atmosphere. Consequently, the climate should not be assumed stationary. The use of EOFs on which to project an observation based upon relationships in a recent well-observed period assumes stationarity and

is therefore invalid. The standard deviation of a linear trend is proportional to the length of the sample record. Therefore any trend pattern will be underestimated in a subsample of the entire record. This accounts for the differences seen in Fig. 6 in which the LDEO values are more suspect. The GISST approach and that of Smith et al. (1996, 1998) separates out the trend before carrying out the remaining analysis and it is clear that such special treatment is essential.

7. Conclusions

Significant differences exist among SST analyses and none is universally the best for all purposes. The previous section discussed the likely origin of the main discrepancies and the need for progress to address them. Relative to the large magnitude of the annual cycle in surface temperature over midlatitude oceans and the large meridional gradient in SST from the Tropics to high latitudes, anomalies in SST are often quite small, yet they can have important impacts on the climate, especially over the Tropics. Accurate climatologies are essential, therefore, in order to monitor climate anomalies and detect climate changes. In the long-term mean climatologies, large differences exist near sea ice at high latitudes and there are differences in spatial resolution of real but small-scale features such as the Gulf Stream and equatorial upwelling, which are better resolved in the NCEP SST analyses. A climatology that resolves these features is necessary to be able to better define and quality control the anomalies when observations are present.

In GISST 2.3b there is evidence that the quality control of observations could be improved and there are major deficiencies in the temporal continuity of SST anomalies after 1981. If existing GISST analyses are used for driving AGCMs, we strongly recommend that a 3- or 5-month running mean filter should be applied to the monthly means after 1981. Satellite data usage in analyses has the advantage of better defining spatial patterns of SST and regions of strong gradient, but more attention needs to be paid to possible biases in retrievals, especially associated with volcanic aerosol, and to physical differences between skin and bulk temperature. Such biases can easily affect trends, and this seems to be the case for the NCEP OI SSTs. Other differences between the SST analyses relate to differences in the processing of in situ data. Differences between the processed in situ data used at NCEP and the MOHSST (version 6) data, for instance,

are larger than differences in the unprocessed in situ data from the two centers (R. W. Reynolds 1999, personal communication), so a careful comparison of the in situ processing methods is warranted.

New methods of interpolating and extrapolating into areas devoid of observations likely improves the SST analyses but caution is called for in what such methods may imply for trends or other nonstationary features of time series. Consequently the long-term trends in the LDEO SST analyses are especially suspect. However, the optimal smoothing analysis technique has much to recommend it, and it is hoped that something like it could be applied to higher-resolution datasets with special measures taken to deal with the trends.

In the course of this study, we have examined ensembles of runs with the CCM3 forced by specified SST sequences from the different analyses. As shown in Fig. 10, there appear to be significant differences between the modeled local precipitation response to the SST anomalies and the estimates from observations. This result has to be tempered by the uncertainty in the observed precipitation estimates and also in the SSTs, as shown here, but it provides a strong indication that such model experiments may not be well posed physically. Implicit in these AGCM experiments is the assumption that the atmosphere responds locally to the SSTs and that the SSTs evolve in a realistic fashion. There is a good basis for the first assumption in the Tropics and subtropics, such as with El Niño events, as the atmosphere responds locally and fairly deterministically (e.g., Shukla 1998; Trenberth et al. 1998). But in the extratropics this is not the case, as more typically it seems that the atmosphere drives the SST changes and the atmospheric response is unlikely to be primarily local (Trenberth et al. 1998). In addition, as chaotic aspects dominate the flow it is unlikely that the SST evolution matches that implied by the surface fluxes in the AGCM simulation. These points add to those of Saravanan (1998), who found that the relationship between the surface heat flux is opposite in sign in AMIP-type integrations versus fully coupled ocean–atmosphere models, so there are limitations to using AGCMs with specified SSTs.

We used the AGCM model results to show that the tropical SST differences that exist in the datasets are important, not only in the region in which they occur but also globally through teleconnections. Differences can be as large as those from modest El Niño events, including the implied differences in temperatures and precipitation. Therefore the SST dataset chosen for

AMIP-type integrations, for attempting to sort out and detect climate signals (e.g., Folland et al. 1998), for monitoring climate, and for global reanalyses matters and should be a factor in evaluating results. We note that in evaluation of the AMIP integrations (Gates et al. 1999), no mention is made of SST uncertainties at all.

Nevertheless, many of the problems can be readily addressed in reprocessing of the SSTs. For GISST, a later version (GISST 3.0) has many of the same characteristics as shown here. But ongoing efforts, for instance, have eliminated the post-1981 lack of temporal continuity by reanalyzing the recent record using an OI technique in place of the Poisson blending technique (N. A. Rayner 1999, personal communication). The result is that the lag-1 month autocorrelations in the reprocessed GISST data [HADISST1 (formerly GISST 4.0), yet to be released] are in much closer agreement with the NCEP OI results shown in Fig. 4.

Other discussions have already occurred among the groups involved with producing the SST analyses, and moves are under way to reanalyze the data in ways that should produce a much improved historical SST record. New analysis techniques at NCEP are reducing (but not entirely eliminating) the cold bias evident in the OI product relative to GISST (Fig. 7), and there appears to be a convergence toward adoption of the UKMO statistical techniques for deriving marginal ice-zone SSTs, although results need to be tested using independent data. Clearly, as the SST analyses evolve, it will be important to continue critical evaluations and comparisons of the datasets.

References

Bottomley, M., C. K. Folland, J. Hsiung, R. E. Newell, and D. E. Parker, 1990: *Global Ocean Surface Temperature Atlas*. U. K. Meteorological Office, 20 pp. and 313 plates.

Cane, M. A., A. C. Clement, A. Kaplan, Y. Kushnir, R. Murtugudde, D. Pozdnyakov, R. Seager, and S. E. Zebiak, 1997: 20th century sea surface temperature trends. *Science*, **275,** 957–960.

Enfield, D. B., and A. M. Mestas-Nuñez, 1999: Multiscale variabilities in global sea surface temperatures and their relationships with tropospheric climate patterns. *J. Climate,* **12,** 2719–2733.

Folland, C. K., and D. E. Parker, 1995: Correction of instrumental biases in historical sea surface temperature data. *Quart. J. Roy. Meteor. Soc.,* **121,** 319–367.

——, R. W. Reynolds, M. Gordon, and D. E. Parker, 1993: A study of six operational sea surface temperature analyses. *J. Climate,* **6,** 96–113.

——, D. M. H. Sexton, D. J. Karoly, C. E. Johnson, D. P. Rowell, and D. E. Parker, 1998: Influences of anthropogenic and oceanic forcing on recent climate change. *Geophys. Res. Lett.,* **25,** 353–356.

Gates, W. L., 1992: AMIP: The Atmospheric Model Intercomparison Project. *Bull. Amer. Meteor. Soc.,* **73,** 1962–1970.

——, and Coauthors, 1999: An overview of the results of the Atmospheric Model Intercomparison Project (AMIP I). *Bull. Amer. Meteor. Soc.,* **80,** 29–55.

Hurrell, J. W., 1996: Influence of variations in extratropical wintertime teleconnections on Northern Hemisphere temperature. *Geophys. Res. Lett.,* **23,** 665–668.

——, J. J. Hack, B. A. Boville, D. L. Williamson, and J. T. Kiehl, 1998: The dynamical simulation of the NCAR Community Climate Model version 3 (CCM3). *J. Climate,* **11,** 1207–1236.

Kaplan, A., Y. Kushnir, M. A. Cane, and M. B. Blumenthal, 1997: Reduced space optimal analysis for historical datasets: 136 years of Atlantic sea surface temperatures. *J. Geophys. Res.,* **102,** 27 835–27 860.

——, M. A. Cane, Y. Kushnir, A. C. Clement, M. B. Blumenthal, and B. Rajagopalan, 1998: Analyses of global sea surface temperature 1856–1991. *J. Geophys. Res.,* **103,** 18 567 –18 589.

Kiehl, J. T., J. J. Hack, G. B. Bonan, B. A. Boville, D. L. Williamson, and P. J. Rasch, 1998: The National Center for Atmospheric Research Community Climate Model: CCM3. *J. Climate,* **11,** 1131–1149.

Knight, R. W., 1984. Introduction to a new sea-ice database. *Ann. Glaciol.,* **5,** 81–84.

Masutani, M., 1997: Relation between SST and rainfall on the seasonal time scale. *Proc. 22d Climate Diagnostics and Prediction Workshop,* Berkeley, CA, Climate Prediction Center, NOAA, 65–68.

Parker, D. E., and C. K. Folland, 1991: Worldwide surface temperature trends since the mid-19th century. *Greenhouse-Gas-Induced Climate Change: A Critical Appraisal of Simulations and Observations,* M. Schlesinger, Ed., Elsevier, 173–193.

——, P. D. Jones, C. K. Folland, and A. Bevan, 1994: Interdecadal changes of surface temperature since the late nineteenth century. *J. Geophys. Res.,* **99,** 14 373–14 399.

——, C. K. Folland, and M. Jackson, 1995a: Marine surface temperature: Observed variations and data requirements. *Climatic Change,* **31,** 559–600.

——, M. Jackson, and E. B. Horton, 1995b: The GISST2.2 sea surface temperature and sea ice climatology. CRTN 63, 16 pp. plus figures. [Available from Hadley Centre for Climate Prediction and Research, Meteorological Office, London Road, Bracknell, Berkshire RG12 2SY, United Kingdom.]

——, C. K. Folland, A. Bevan, M. N. Ward, M. Jackson, and K. Maskell, 1995c: Marine surface data for analysis of climatic fluctuations on interannual to century timescales. *Natural Climate Variability on Decade-to-Century Time Scales,* D. G. Martinson et al., Eds., National Academy Press, 241–250.

Rayner, N. A., E. B. Horton, D. E. Parker, C. K. Folland, and R. B. Hackett, 1996: Version 2.2 of the Global Sea-Ice and Sea Surface Temperature data set, 1903–1994. CRTN 74, 21 pp. plus figures. [Available from Hadley Centre for Climate Prediction and Research, Meteorological Office, London Road, Bracknell, Berkshire, RG12 2SY, United Kingdom.]

Reynolds, R. W., 1988: A real-time global sea surface temperature analysis. *J. Climate,* **1,** 75–86.

——, and D. C. Marsico, 1993: An improved real-time global SST analysis. *J. Climate,* **6,** 114–119.

——, and T. M. Smith, 1994: Improved global sea surface temperature analyses using optimum interpolation. *J. Climate,* **7,** 929–948.

——, and ——, 1995: A high-resolution global sea surface temperature climatology. *J. Climate,* **8,** 1571–1583.

Saravanan, R., 1998: Atmospheric low-frequency variability and its relationship to midlatitude SST variability: Studies using the NCAR Climate System Model. *J. Climate,* **11,** 1386–1404.

Shukla, J., 1998: Predictability in the midst of chaos: A scientific basis for climate forecasting. *Science,* **282,** 728–731.

Smith, T. M., and R. W. Reynolds, 1998: A high-resolution global sea surface temperature climatology for the 1961–90 base period. *J. Climate,* **11,** 3320–3323.

——, ——, R. E. Livezey, and D. C. Stokes, 1996: Reconstruction of historical sea surface temperatures using empirical orthogonal functions. *J. Climate,* **9,** 1403–1420.

——, R. E. Livezey, and S. S. Shen, 1998: An improved method for analyzing sparse and irregularly distributed SST data on a regular grid: The tropical Pacific Ocean. *J. Climate,* **11,** 1717–1729.

Trenberth, K. E., and A. Solomon, 1993: Implications of global atmospheric spatial spectra for processing and displaying data. *J. Climate,* **6,** 531–545.

——, and T. J. Hoar, 1996: The 1990–1995 El Niño–Southern Oscillation event: Longest on record. *Geophys. Res. Lett.,* **23,** 57–60.

——, and ——, 1997: El Niño and climate change. *Geophys. Res. Lett.,* **24,** 3057–3060.

——, J. R. Christy, and J. W. Hurrell, 1992: Monitoring global monthly mean surface temperatures. *J. Climate,* **5,** 1405–1423.

——, G. W. Branstator, D. Karoly, A. Kumar, N.-C. Lau, and C. Ropelewski, 1998: Progress during TOGA in understanding and modeling global teleconnections associated with tropical sea surface temperatures. *J. Geophys. Res.,* **103,** 14 291–14 324.

Walsh, J. E., 1995: A sea ice database. *Workshop on Simulations of the Climate of the Twentieth Century using GISST,* C. K. Folland and D. P. Rowell, Eds., CRTN, 56, 54–55.

Webster, P. J., C. A. Clayson, and J. A. Curry, 1996: Clouds, radiation and the diurnal cycle of sea surface temperature in the tropical western Pacific Ocean. *J. Climate,* **9,** 1712–1730.

Woodruff, S. D., R. J. Slutz, R. L. Jenne, and P. M. Steurer, 1987: A Comprehensive Ocean–Atmosphere Data Set. *Bull. Amer. Meteor. Soc.,* **68,** 1239–1250.

Xie, P., and P. A. Arkin, 1996: Analyses of global monthly precipitation using gauge observations, satellite estimates, and numerical model predictions. *J. Climate,* **9,** 840–858.