The Effect of Aerosols and Clouds on the Retrieval of Infrared Sea Surface Temperatures

JORGE VÁZQUEZ-CUERVO AND EDWARD M. ARMSTRONG

JPL, California Institute of Technology, Pasadena, California

ANDY HARRIS

NOAA Cooperative Institute for Climate Studies, Earth System Science Interdisciplinary Center, University of Maryland, College Park, College Park, Maryland

(Manuscript received 19 May 2003, in final form 7 April 2004)

ABSTRACT

Comparisons are performed between spatially averaged sea surface temperatures (ASST2) as derived from the second Along-Track Scanning Radiometer (ATSR-2) on board the second European Remote Sensing Satellite (*ERS-2*) and the NOAA–NASA Advanced Very High Resolution Radiometer (AVHRR) Oceans Pathfinder dataset (MPFSST). Difference maps, MPFSST − ASST2, along with the application of a simple statistical regression model to aerosol and cloud data from the Total Ozone Mapping Spectrometer (TOMS), are used to examine the impact of possible aerosol and cloud contamination. Differences varied regionally, but the largest biases were seen off western Africa. Nighttime and daytime differences off western Africa were reduced from −0.5° to −0.2°C and from −0.1° to 0°C, respectively. Significant cloud flagging, based on the model, occurred in the Indian Ocean, the equatorial Pacific, and in the vicinity of the Gulf Stream. Comparisons of the MPFSST and the ASST2 with in situ data from the 2002 version of the World Oceanic Database (WOD02) off western Africa show larger mean differences for the MPFSST. The smallest mean differences occurred for nighttime ASST2 − WOD02 with a value of 0.0° ± 0.4°C.

1. Introduction

Previous studies have indicated that cloud contamination (e.g., Jones et al. 1996) and aerosols (e.g., May et al. 1992) are sources of error in the retrieval of satellite infrared derived sea surface temperatures (SSTs). Results from Simpson et al. (2001) focus on improving algorithms for better removal of residual cloud cover. Studies of climate change require that these errors be understood and removed before accuracies of 0.1°C can be achieved (e.g., Merchant and Harris 1999). Additionally, the assimilation of SSTs into general circulation models necessitates that error sources be quantified. Attempts at future merging of satellite-derived SST datasets (e.g., Donlon 2002) using optimal interpolation and assimilation approaches require an accurate knowledge of error sources from regional to global scales. This paper is an attempt to quantify the spatial and temporal characteristics of these errors, specifically as they relate to contamination of SST retrievals by aerosols and undetected clouds.

Corresponding author address: Dr. Jorge Vázquez-Cuervo, JPL, California Institute of Technology, M/S 300/323, 4800 Oak Grove Dr., Pasadena, CA 91109.
E-mail: jv@pacific.jpl.nasa.gov

To accomplish this the study incorporates the use of several datasets, including the modified Pathfinder sea surface temperature (MPFSST) dataset, averaged global sea surface temperatures (ASST2) from the second Along-Track Scanning Radiometer (ATSR-2) on board the second European Remote Sensing Satellite (*ERS-2*), and aerosol and cloud reflectivity data from the Total Ozone Measuring Spectrometer (TOMS). Although the ASST2 is a direct measure of the skin temperature, while the MPFSST is a measured skin temperature tuned toward a bulk temperature (Kilpatrick et al. 2001), correlations of statistical significance with possible error sources such as aerosols would indicate areas where problems might remain in the infrared derived SSTs. Because both aerosols and clouds can have an effect on the measured SST, correlations are better done using the differences (MPFSST − ASST2). The rationale for the correlations done on the differences is that, individually, both aerosols and clouds could be correlated with SST. Additionally, Saharan dust is carried by the same trade winds that cause the cold upwelling off the west coast of Africa. Aerosols can block incoming solar radiation, reducing both bulk and skin measured SST, while the formation of clouds is well known to be influenced by SST and surface fluxes. By performing the correlations

on the differences, this covariability between the MPFSST, ASST2, and aerosol influenced SSTs is removed. Of course, this would not remove any differences that remain as the result of the bulk (MPFSST) versus (ASST2) skin measured SST. An example might be surface fluxes driven by skin temperatures that affect cloud formation. The analysis of the skin versus bulk temperature issue is beyond the scope of this paper and will be left for future research. A rationale for the continued study of the MPFSST and ASST differences is seen from earlier comparison results.

Earlier comparisons between the ATSR-1 derived SSTs (ASST1) and the MPFSST (Vazquez-Cuervo and Sumagaysay 2001, hereafter referred to as VCS) indicated extensive problems with cloud contamination in the ASST1 datasets. Mean differences for both the daytime and nighttime retrievals were found to be greater than 1.3° ± 0.6°C. After application of a semiannual and annual harmonic model (Jones et al. 1996) to the data the mean differences for the nighttime data were reduced from 1.6° ± 0.8°C to approximately 1.4° ± 0.6°C. The application of a similar model to the daytime data did not reduce the mean difference or standard deviation significantly—an expected result since the model was fitted to the daytime data. These results led to the conclusion that a significant percentage of the mean differences between the MPFSST and ASST1 nighttime data were due to cloud contamination. Note, however, that the regional impacts were much larger, particularly in areas where marine stratiform cloud is seasonally prevalent. Such comparison work is deemed important because the MPFSST and ASST datasets strive to be of climate quality and may both be incorporated in future SST analyses once their quality has been demonstrated and sources of residual bias understood, quantified, and removed.

Merchant and Harris (1999) showed that the biases with respect to in situ data which existed in the original ATSR-1 SST data were substantially reduced with the application of aerosol robust SST retrieval coefficients derived using a new water vapor continuum parameterization in the radiative transfer model. Similar SST retrieval coefficients also have been developed and applied lately to ATSR-2 data. This work focuses on comparisons between the ASST2 and the MPFSST in order to determine the effect of these improved retrieval coefficients on MPFSST − ASST2 differences, and external datasets are utilized to assist in diagnosis of the underlying causes of remaining biases.

Cloud and aerosol data from the Total Ozone Mapping Spectrometer (TOMS; Chiapello et al. 2000; McPeters et al. 1998) are used to develop a simple linear model and flagging methodology for residual cloud and/or aerosol contamination in the MPFSST and ASST2 fields. The assumption in the methodology used will be that significant correlation between the MPFSST − ASST2 and the TOMS aerosol and cloud data is due to residual cloud and/or aerosol contamination in either the MPFSST or ASST2 data. Significant correlations at the 99% level of significance were defined at 0.2. This was based on 147 degrees of freedom. Other causes affecting the MPFSST − ASST2 differences are atmospheric water vapor, and the skin–bulk temperature variability due to methodologies of calculating the MPFSST and the ASST2.

Values of MPFSST − ASST2 are then compared before and after the application of the flagging based on the linear regression between the MPFSST − ASST2 values and the TOMS aerosol and cloud data (see next section). The flagging is defined as the exclusion of those values having significant correlations and should result in reduction of differences between the MPFSST and ASST2. Any reduction in the mean difference is a good indication of cloud and/or aerosol contamination in either the ASST2 or the MPFSST. Thus, these difference maps, along with Hovmoeller plots, provide a basis for examining and quantifying error sources in the calculation of SST. The addition of the TOMS data provides an independent technique for identifying cloud or aerosol contaminated SST values.

Following this introduction, the paper is divided into eight sections. Section 2 will summarize the important dataset details and the gridding. Only summaries will be given because of the similarity with the methodology used in VCS. Section 3 will focus on results and statistics from time versus latitude plots of MPFSST − ASST2. Section 4 will highlight some results taken from regional differences in the equatorial Pacific, North Atlantic, and other ocean basins. Section 5 will show results from the correlation analysis with the TOMS aerosol and cloud data, while section 6 focuses on results before and after applying the flagging scheme to the MPFSST and the ASST2. Section 7 concentrates on examining regional differences off the African coast. Section 8, as a discussion, offers possible explanations for areas of the large MPFSST − ASST2, followed by overall conclusions in section 9.

2. Data

Four different datasets were used in the study, covering the period of time from July 1996 to June 1999. These are the MPFSST, ASST2, and the TOMS aerosol and cloud products. A brief description will be given for each dataset, with references where more details may be found.

The Pathfinder Advanced Very High Resolution Radiation (AVHRR) Oceans SST data used in this work were done using version 4.1 of the processing algorithms (Kilpatrick et al. 2001). The SST retrieval algorithm is based on the nonlinear formulation of Walton (1988):

$$\mathrm{SST} = a_1 + a_2 T_4 + a_3 (T_4 - T_5) T_{\mathrm{surf}} + a_4 (\sec\varphi - 1)(T_4 - T_5),$$

where a_1, a_2, a_3, a_4 are coefficients based on a least squares fit to in situ data and T_4, T_5 are the brightness temperatures in channels 4 and 5, corresponding to center wavelengths of ~11 and ~12 μm, respectively. The satellite scan angle is φ and T_{surf} is a first-guess sea surface temperature field; in this case supplied from the Reynolds and Smith (1994) optimally interpolated sea surface temperature analysis. In the MPFSST, coefficients are derived on a monthly basis by fitting to 5 months of satellite—in situ matchup data centered on the month in question, and separate coefficients are derived for ''low'' and ''intermediate to high'' water vapor burdens using the split-window channel difference as a proxy. Thus, this form of the algorithm, known as the MPFSST, calculates a skin temperature that is tuned toward a bulk temperature. The MPFSST data are available at several temporal and spatial resolutions with comparisons in this paper done using the daily 54-km version 4.1 best pixel data. The daily 54-km files were then binned into weekly files. The daily files were used in the weekly binning and not the 8-day MPFSST weekly distributable product because of the desire to achieve an exact beginning and ending collocation time of the MPFSST with the ASST2. Thus, comparisons were done between weekly binned files, but the daily MPFSST and ASST2 files were used in the binning to ensure that the beginning and ending collocation times of the MPFSST and ASST2 were aligned. This also allowed for the comparison of separate daytime and nighttime fields, although the different overpass times may cause problems as the diurnal cycle operates on shorter time scales than 12 h.

The ASST2 data (ASST) are available directly from the Rutherford Appleton Laboratory Web site (online at http://www.atsr.rl.ac.uk) from 4 May 1995 through the present with a gap of several months in 1996 due to instrument malfunction. The original ASST2 (corresponding to v321 of the processing software) 18-km data were binned and gridded into weekly, 0.5° files. This gridding and binning allowed the data to be collocated in space and time with the MPFSST and the TOMS data. Details of the SST retrieval algorithm have been published recently in Merchant et al. (1999).

The ATSR-2 instrument has three bands (3.7, 11, and 12 μm) in the infrared and one band in the near-infrared (1.6 μm) range. Additionally three other bands are available in the visible range. A major advantage of the ATSR-2 instrument over the ATSR-1 is the continued availability of the 3.7-μm channel, which failed on the ATSR-1 instrument less than a year into the mission. The dual view capability of the ATSR 1-2 instrument allows for a greater inherent accuracy in the SST calculation (e.g., Murray et al. 1996). SSTs from the ATSR instrument are derived from a combination of observed brightness temperatures in the thermal infrared:

$$\text{SST} = a_0 + \sum a_i T_i,$$

where a_i are coefficients statistically determined from top-of-atmosphere channel brightness temperatures (T_i) simulated using a radiative transfer model and representative atmospheric and surface data (e.g., Závody et al. 1995; Merchant et al. 1999), instead of direct regression of satellite-observed brightness temperatures to SST in situ observations, as in the MPFSST.

The global aerosol data used in this study was derived from the TOMS instrument. (Information on this instrument may be found under: http://toms.gsfc.nasa.gov/.) From 25 July 1996 to 28 June 1997 data were retrieved from the instrument on board Japan's *Advanced Earth Observing Satellite* (*ADEOS*) spacecraft, which failed on 30 June 1997. Data from 25 July 1996 to the present are also available from the TOMS instrument on board the National Aeronautics and Space Administration (NASA) *Earth Probe* satellite and therefore forms the basis of this work. The algorithm is based on the scattering in the ultraviolet bands at 331 and 360 nm (Chiapello et al. 2000). Positive values of the aerosol index (AI) denote ultraviolet absorbing aerosols while negative values denote nonabsorbing aerosols. Briefly, the AI is based on the scattered radiance compared to a pure atmosphere. It is defined as the ratio of the radiance between the two spectral bands at 331 and 360 nm. The AI values are available globally in daily gridded maps at a spatial resolution of 1° latitude and 1.25° longitude. Reflectivity, a measure of cloud cover, is also available at the same spatial and temporal resolutions. These datasets were binned to the same weekly 0.5° space–time grids as the ASST2 and the MPFSST to allow direct comparisons between the four datasets. For more information on the processing and algorithm of these datasets, see Chiapello et al. (2000) and McPeters et al. (1998).

3. Global comparisons

Figure 1 shows the time versus latitude plots of MPFSST − ASST2 zonal mean differences from July 1996 to October 1999. Positive daytime values (MPFSST warmer than ASST2) were associated with large differences in the Northern Hemisphere that have a pronounced annual cycle. Nighttime differences show a different trend with MPFSST cooler than ASST2 dominating the Northern Hemisphere. Table 1 summarizes the global statistics of these differences and compares them with previous results from the MPFSST − ASST1 (VCS).

The differences between the ASST1 and ASST2 statistics reflect the application of new coefficients as well as the availability of the 3.7-μm channel, which enabled the suboptimal (i.e., pre-Merchant et al. 1999) methodology employed to develop the initial SST retrieval coefficients to obtain satisfactory results. Because the MPFSST has not changed, any reductions in the mean differences and standard deviations are primarily due to differences in the processing methodology (and instrument capabilities) for ASST1 and ASST2. However,

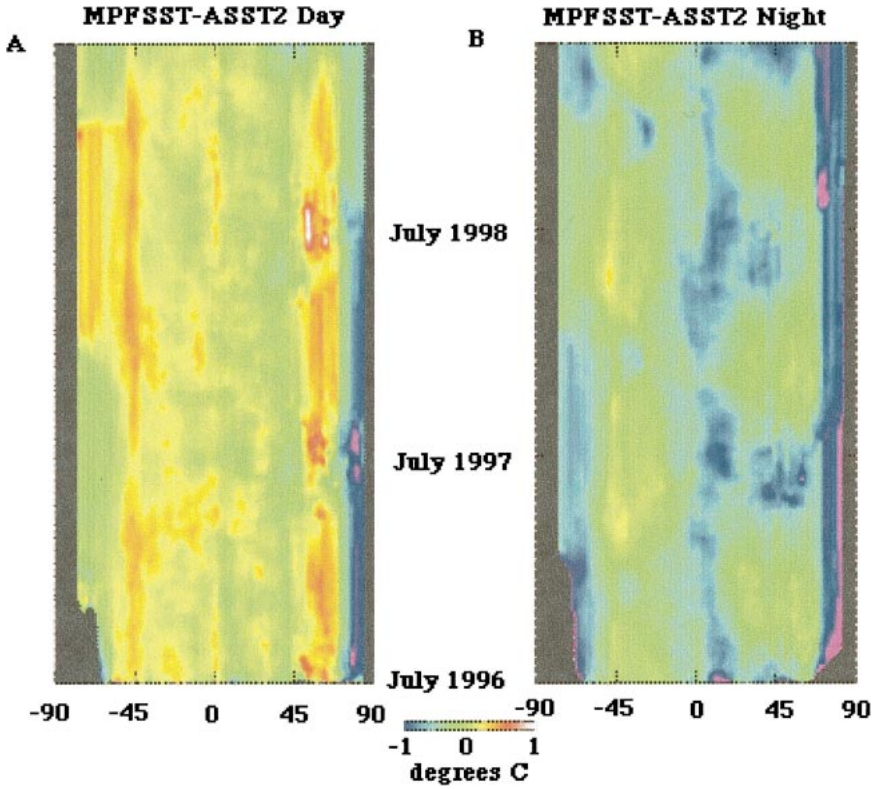

FIG. 1. (a) Time vs latitude mean daytime difference maps of MPFSST − ASST2 (°C). Color scale ranges from −1° to 1°C with minimum values of −1°C (purple) and maximum values of 1°C (red to white). Values were averaged meridionally to derive the mean difference for the given zonal or latitudinal band. (b) Same as (a) except for nighttime values.

remaining differences in regions of high aerosols need to be examined separately for the MPFSST and the ASST2 to determine their effect. The next question becomes whether these global statistics vary and if so what are the regional/basinwide statistics, which are examined in the next section.

4. Regional statistics

Six different ocean basins were examined for calculating regional differences of the MPFSST − ASST2. These regions were selected in an attempt to represent the typical range of SST retrieval conditions likely to be encountered.

Figures 2a–f show the time series of the mean differences for the following basins: the North Atlantic, Mediterranean, Caribbean, western and eastern equatorial Pacific, and the Indian Ocean. In all cases daytime means were positive, indicating the MPFSST was warmer than the ASST2; while nighttime means were negative. The western North Atlantic, Mediterranean, and Caribbean are shown in Figs. 2a–c while the western and eastern equatorial Pacific and the Indian Ocean are shown in Figs. 2d–f. Daytime mean differences were generally greater than the nighttime differences. The largest differences occurred during the daytime in the Mediterranean and the Caribbean. However, the nighttime differences in both basins approached zero. The largest mean nighttime differences occurred in the western North Atlantic and the western equatorial Pacific. Error bars are the largest in the North Atlantic, possibly associated with the higher variability of the Western Boundary Current. The possibility that this is due to cloud contamination will be examined in the following sections. Whether such differences between the nighttime and daytime fields can be explained by the diurnal cycle, or a skin–bulk temperature effect, should also be investigated but is beyond the scope of this present work.

5. MPFSST − ASST2 correlations with aerosols and clouds

Correlations between the MPFSST − ASST2 and the aerosol and cloud data from the TOMS instrument are calculated to identify the regional areas where possible errors still exist. The daytime and nighttime correlations between the MPFSST − ASST2 and the TOMS cloud data (not shown) show statistically significant correlation only in the equatorial regions and in the Indian Ocean, while the majority of the earth is not significant at the 99% level. The negative correlation indicates that

TABLE 1. Global statistics of MPFSST − ASST2 and MPFSST − ASST1 (Vazquez-Cuervo and Sumagaysay 2001).

Pass/comparison	Mean differences (°C)	Std dev (°C)
Daytime (MPFSST − ASST2)	0.25	0.27
Daytime (MPFSST − ASST1)	1.31	0.63
Nighttime (MPFSST − ASST2)	−0.15	0.29
NIghttime (MPFSST − ASST1)	1.56	0.76

during times of cloud cover the MPFSST is lower than the ASST2. One possible explanation would be cloud contamination in the MPFSST. This would be consistent with clouds being cooler than the actual SST. However, it should be borne in mind that there is the potential for cloud contamination in the forward view of the ATSR to bias the retrieved SST warm because the weights applied to brightness temperatures from those channels are negative. Since the projected instrument field of view is approximately 4 times larger in the forward view, subpixel cloud may be both more prevalent and harder to detect.

Figures 3a,b shows the daytime and nighttime correlation of the MPFSST − ASST2 with the aerosol data from TOMS. The largest signals appear off the western coast of Africa. Both the daytime and nighttime images show statistically significant negative correlations in known areas of high aerosol content associated with dust storms (Chiapello et al. 2000). The negative sign of the correlation, in the MPFSST − ASST2, is consistent with cooler MPFSST values being retrieved during times of high aerosol content indicating that perhaps absorbing aerosols are having a greater effect on the retrieval of the MPFSST than ASST2. The following section concentrates on how these results are used to flag clouds and/or aerosols in the MPFSST and the ASST2.

6. Modeling effects of clouds and aerosols on satellite SST retrievals

The assumption in the methodology is that statistically significant correlation between the differences as defined by MPFSST − ASST2 and the TOMS aerosol and cloud data indicate that, in a linear sense, clouds and aerosols are correlated with the variability not common to both the MPFSST and the ASST2. Thus, the goal is to determine those areas where the MPFSST − ASST2 variability is linearly dependent on aerosols and/or clouds and to develop a methodology to flag those values based on the statistical relationship.

A linear regression between MPFSST − ASST2 and the aerosols and clouds was performed at each location (x_i, y_j) where x_i represents longitude and y_j latitude for $i = 1$ to 720 and $j = 1$ to 360, corresponding to 54-km resolution. The two linear regressions were applied at each point (x_i, y_j); MPFSST − ASST2 versus TOMS aerosol and cloud indices. Thus, the regressions were applied to each cell at the binned 54-km spatial and weekly temporal resolutions for 147 successive gridded maps of MPFSST − ASST2, clouds, and aerosols from July 1996 through May 1999. Only regressed data with correlations higher than 0.2 (statistical significance at the 99% level of confidence for 147 degrees of freedom) were retained. The best fit of the regression is then used as a predictor of the bias for a given aerosol and/or cloud value. All values with a predicted absolute difference greater than 0.1°C are flagged as cloud and/or aerosol contaminated. The value of 0.1°C was chosen based on the estimated accuracy of SST needed for climate studies (Allen et al. 1994). In the area off the western coast of Africa, where correlations are statistically significant, the error associated with the best fits to the MPFSST − ASST2 versus aerosols at individual (x, y) locations was 0.2°C. Thus, because of the scatter in the application of the linear model at each (x, y), our flagging is accurate to $0.1° \pm 0.2$°C. After the flagging a new set of MPFSST − ASST2 differences are calculated, with the flag values removed.

As an example at a given location (x_i, y_j) a linear regression is performed between the MPFSST − ASST2 and the cloud and aerosol TOMS data. An $(\text{MPFSST} - \text{ASST2})_{\text{aer}}$, and $(\text{MPFSST} - \text{ASST2})_{\text{clo}}$ are calculated based on the regression:

$$\text{SSTDIF}_{\text{aer}}(x_i, y_j, t) = m_{\text{aer}}(x_i, y_j)[\text{AI}(x_i, y_j, t) + b_{\text{aer}}(x_i, y_j)], \quad (1)$$

$$\text{SSTDIF}_{\text{clo}}(x_i, y_j, t) = m_{\text{clo}}(x_i, y_j)[\text{CI}(x_i, y_j, t) + b_{\text{clo}}(x_i, y_j)], \quad (2)$$

where SSTDIF represents the difference as defined by MPFSST − ASST2, $\text{SSTDIF}_{\text{aer}}$ and $\text{SSTDIF}_{\text{clo}}$ are the best fits to the aerosol and cloud data, m and b, respectively, represent the slope and intercept and $\text{AI}(x_i, y_j, t)$, $\text{CI}(x_i, y_j, t)$ are the aerosol and cloud indices, respectively, for the given location and time, t. Here, x_i and y_j vary from −89.75° to 89.75°N, and −179.75° to 179.75°W, respectively, resulting in 360° latitude by 720° longitude locations for which the regression is performed. This procedure is repeated separately for the daytime and nighttime fields.

For locations (x_i, y_j) where the absolute value of the correlation between the SSTDIF and the aerosol and cloud data is greater than 0.2, the best fit $\text{SSTDIF}_{\text{aer}}$ and $\text{SSTDIF}_{\text{clo}}$ is used as a predictor of the bias between the MPFSST and the ASST2. If the absolute value of the bias is greater than 0.1°C, the value of SSTDIF is flagged as aerosol or cloud contaminated. Once the flagging has been applied a new set of differences noted as $\text{SSTDIF}_{\text{naer}}(x_i, y_j, t)$ and $\text{SSTDIF}_{\text{nclo}}(x_i, y_j, t)$ can be calculated. In cases where both the aerosol and cloud flagging were applied, a new $\text{SSTDIF}_{\text{nflag}}(x_i, y_j, t)$ is defined. The object at this stage is not to apply a correction scheme for aerosols or clouds but only to determine whether applying a simple flagging mechanism could reduce the differences between the two datasets, while

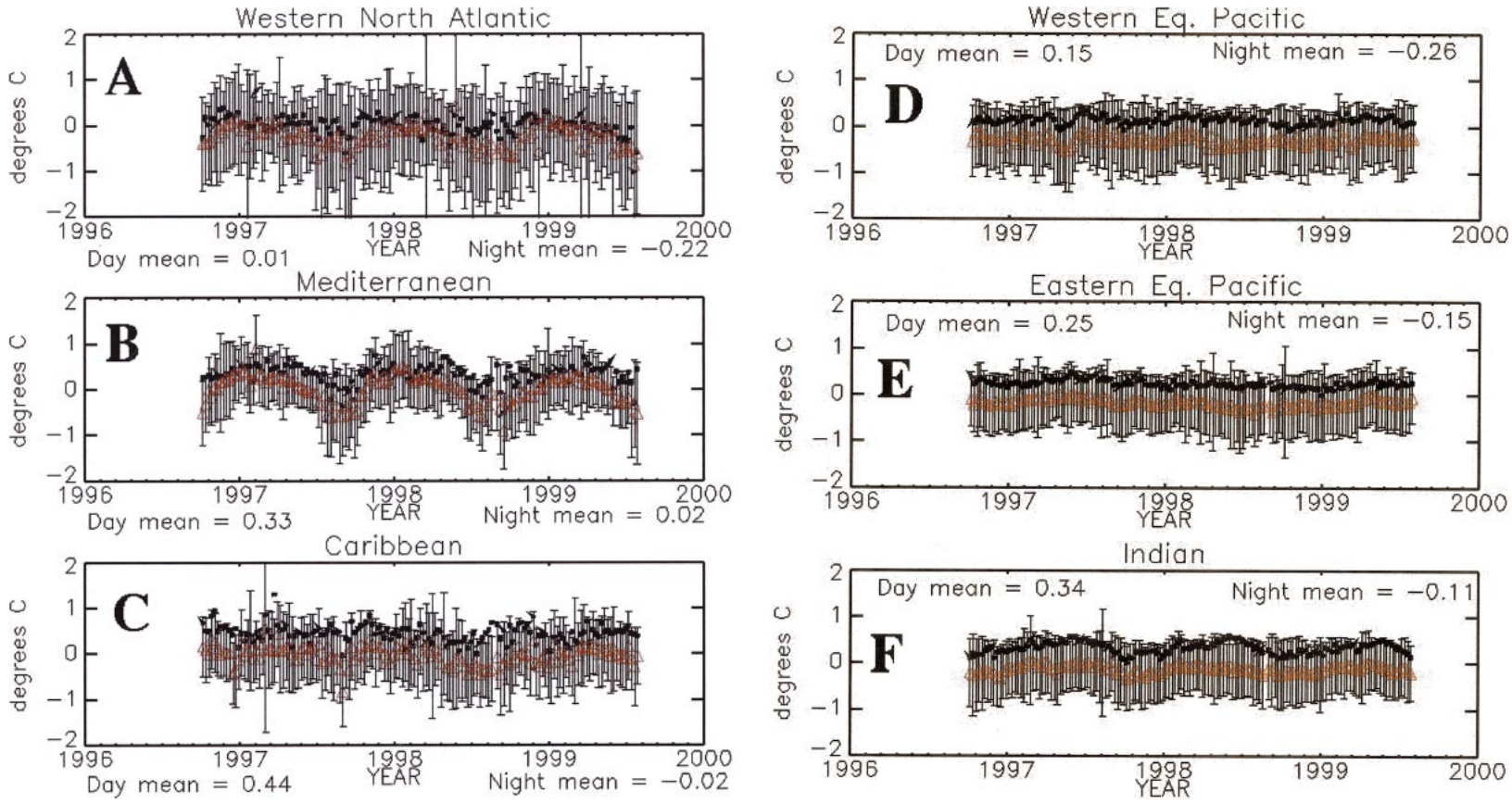

FIG. 2. (a) Time series of spatially averaged MPFSST − ASST2 differences for the western North Atlantic. Error bars are for 1 std dev. The *y* axis ranges from −1° to 1°C and *x* axis from Jul 1996 to May 1999. Red triangles indicate nighttime means, while black indicates daytime means. (b)–(f) Same as (a) except for (b) Mediterranean, (c) Caribbean, (d) western equatorial Pacific, (e) eastern equatorial Pacific, and (f) Indian Ocean.

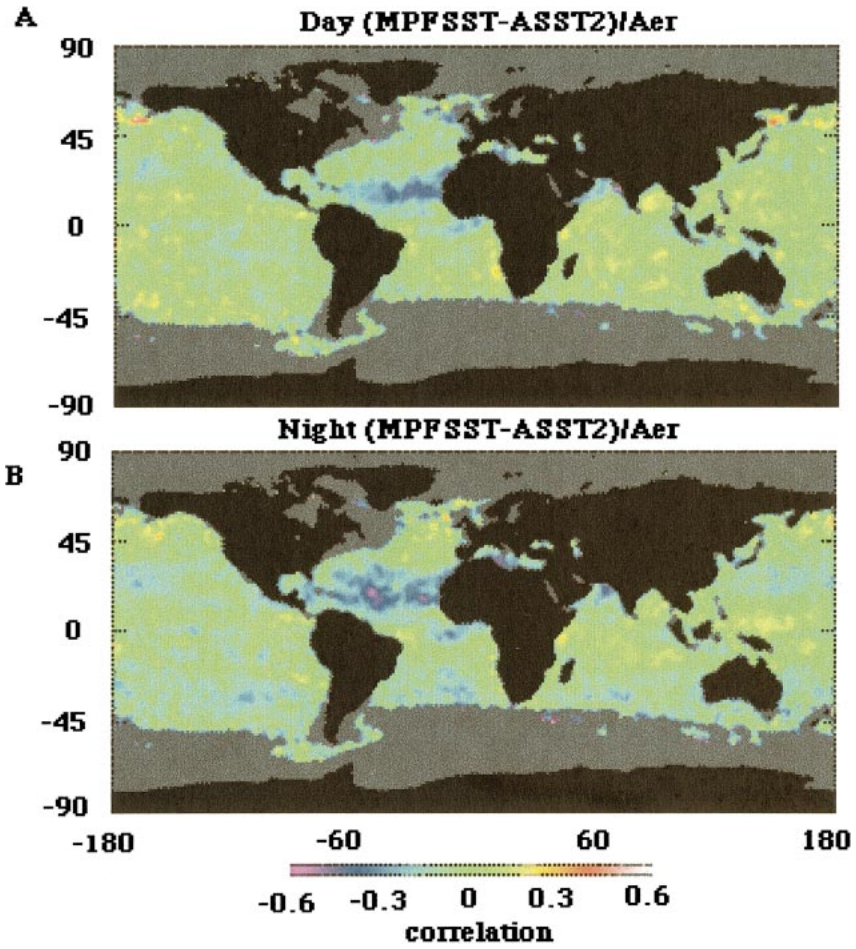

FIG. 3. (a) Correlation of daytime MPFSST − ASST2 values with aerosol data from the TOMS instrument. Color scale ranges from −0.6 to 0.6 with minimum values in purple and maximum values in red and white. Gray values correspond to areas where collocated MPFSST − ASST2 and cloud data were not sufficient to calculate correlation. (b) Same as (a) except for nighttime MPFSST − ASST2.

additionally identifying those areas where aerosol or cloud contamination still exists. This is considered as a first step before any future attempt to apply a correction is attempted. The application of the correlation to the MPFSST − ASST2, instead of the MPFSST, ASST2 individually, makes it problematic for applying a correction since the differences do not allow for identification of the error in the MPFSST and/or the ASST2. However, comparisons with in situ data (see section 8) indicate that the ASST2 is doing better than the MPFSST at correcting for aerosols.

TABLE 2. SSTDIF and $SSTDIF_{flag}$.

Pass/comparison	Mean differences (°C)	Std dev (°C)
Daytime	0.17	0.25
Nighttime	−0.15	0.24
Daytime/flag	0.14	0.24
Nighttime/flag	−0.12	0.23

The latitude versus time values of the daytime results (not shown) for the SSTDIF and the $SSTDIF_{nflag}$, where $SSTDIF_{nflag}$ are the difference maps with both the aerosol and cloud flagging applied show little effect from the flagging. For the nighttime data, prior to flagging, the largest negative differences are found at approximately 10°N with a clear annual cycle and large negative differences found in the later summer to early fall time frame. The aerosol flagging results in a significant reduction in the mean nighttime differences, especially at 10°N. Table 2 summarizes the global statistics of the SSTDIF and the $SSTDIF_{nflag}$. Global mean differences were calculated between 50°S and 50°N to avoid biasing the calculation due to noise in the fitting in the high latitudes.

Although global mean differences were reduced after the application of the flags for both the daytime and nighttime fields, the reduction was only on the order of hundredths of a degree. Global and even zonal averages

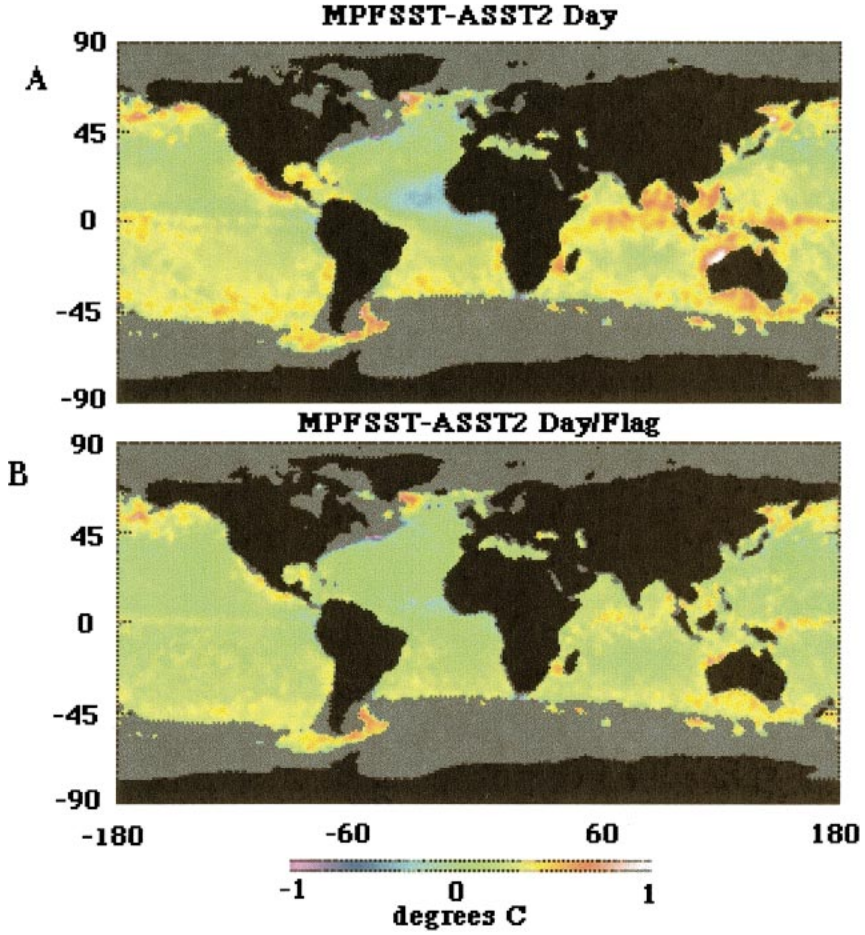

FIG. 4. (a) Temporally averaged spatial map of daytime MPFSST − ASST2. Color scale ranges from −1° to 1°C with minimum value of −1°C (purple) and maximum values of 1°C (red to white). (b) Same as (a) except with aerosol and cloud flagging.

are clearly not enough to identify regions where aerosol and residual cloud contamination may still exist.

7. Regional trends of SSTDIF and $SSTDIF_{flag}$

Figures 4a,b and 5a,b show the temporally averaged spatial map of the mean differences of SSTDIF and $SSTDIF_{nflag}$ for the daytime and nighttime fields. Daytime differences of SSTDIF prior to flagging are dominated by positive values in the northern Indian Ocean and the Southern Ocean, and negative values off the West African coast in the equatorial Atlantic. Both the positive mean differences in the Indian Ocean and the negative differences off the western coast of Africa are reduced significantly by the flagging procedure.

Nighttime differences (Fig. 5a) are dominated by large negative values off the western coast of Africa and along the eastern equatorial Atlantic. Figure 5b shows a dramatic reduction in these differences upon application of the aerosol and cloud flagging.

To determine the number of flagged aerosol values, Figs. 6a,b show the percentage of SSTDIF values flagged as aerosol. Figures 6a,b show that off the coast of Africa up to 60% of the pixel values were flagged as aerosol for both the daytime and nighttime fields. Values over the rest of the global ocean are generally lower than 10%. Values for cloud flagging (not shown) indicate the maximum exists during the nighttime for the Arabian Sea and the Bay of Bengal where more than 60% of the pixels were flagged as cloud.

The area with the largest signal in the SSTDIF is off the western coast of Africa in the equatorial Atlantic. Here the daytime and nighttime fields showed a significant reduction in the $SSTDIF_{naer}$ from the SSTDIF. The daytime differences for SSTDIF and $SSTDIF_{naer}$ were reduced from $-0.10° \pm 0.16$°C to $0.0° \pm 0.13$°C. A more dramatic reduction occurred in the nighttime values, which were reduced from $-0.53° \pm 0.21$°C to $-0.22° \pm 0.16$°C. These results indicate that the flagging is effective in reducing the mean differences and standard deviations of the SSTDIF. The next section will focus on the large SSTDIF signal off the western coast of Africa and explore some of the reasons

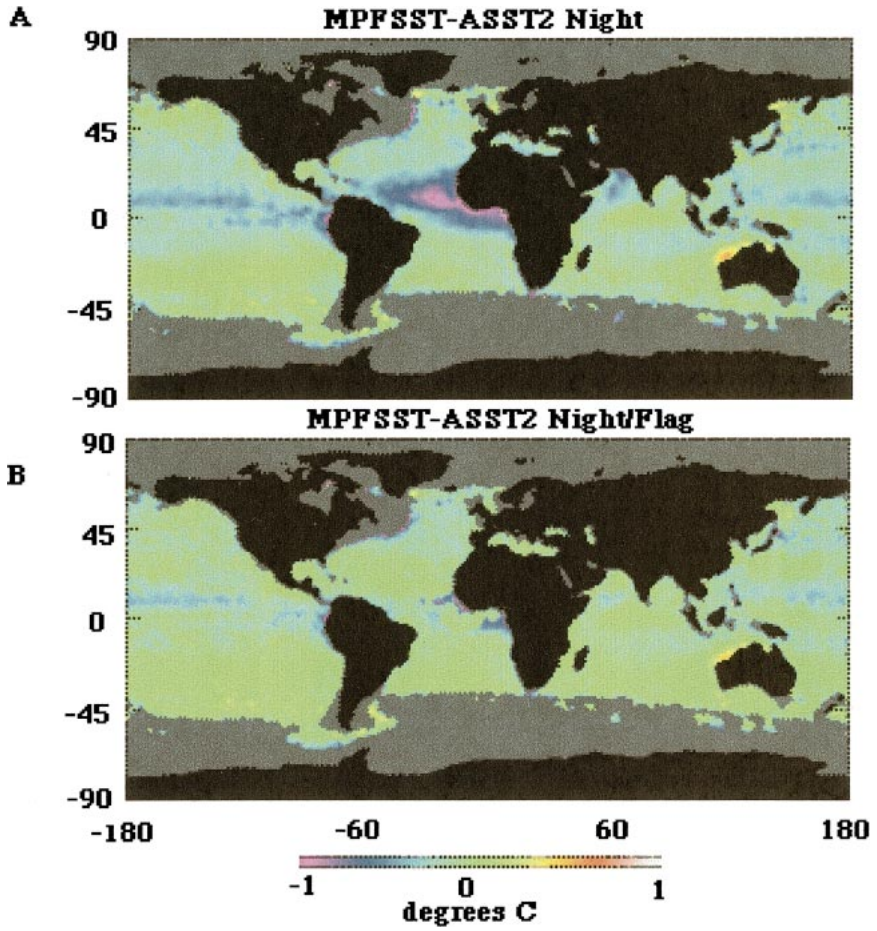

FIG. 5. Same as in Fig. 4 except for nighttime MPFSST − ASST2.

for the mean differences and their reductions via flagging.

8. Discussion

To further confirm the effect of aerosols off the western coast of Africa, an examination is made of the consistency of the phase between the aerosol index (AI) (TOMS) and the Pathfinder Atmosphere project (PATMOS) aerosol data.

Figure 7 shows the plot of the spatially averaged (AI) along the West African coast from both the TOMS satellite and the AVHRR 1-channel-derived algorithm. The AVHRR 1-channel algorithm is shown in red and covers a limited period in 1998. The PATMOS aerosol data are available through an FTP site (ftp://aries.nesdis.noaa.gov). Both datasets show a pronounced semiannual signal with maxima in the late winter and midsummer time frames, consistent with the phasing of SSTDIF in that region. Mean nighttime differences on the order of −0.5°C that are reduced to −0.2°C after the application of the aerosol flagging, and the consistency of the SSTDIF (see Fig. 1) with the seasonal Saharan dust signal, are strong indicators that aerosols play a significant part in affecting the retrieval of SSTs off the West African coast.

Further indication of the aerosol effect on the calculation of the MPFSST and the ASST2 can be seen in independent comparisons of satellite SST with in situ data from the World Ocean Database 02 (WOD02; Conkright et al. 2002). Matchups were collocated within a 6-h time window. Comparisons were carried out using MPFSST retrievals based on quality flags of 4 and above and flag 7 only (7 is the highest quality Pathfinder SST flag) (Kilpatrick et al. 2001). Table 3 summarizes the results of these comparisons and illustrates that the ASST2, from a statistical analysis based on the mean differences, is making a more accurate SST measurement than MPFSST off the West African coast. ASST2 nighttime comparisons with in situ data are impressive in showing a mean difference of $-0.03° \pm 0.41$°C.

Although specific regional differences have been shown to be dependent on cloud and aerosol contamination, it is possible that a significant portion of the global differences between the MPFSST and the ASST2 can be attributed to the diurnal cycle and/or skin–bulk

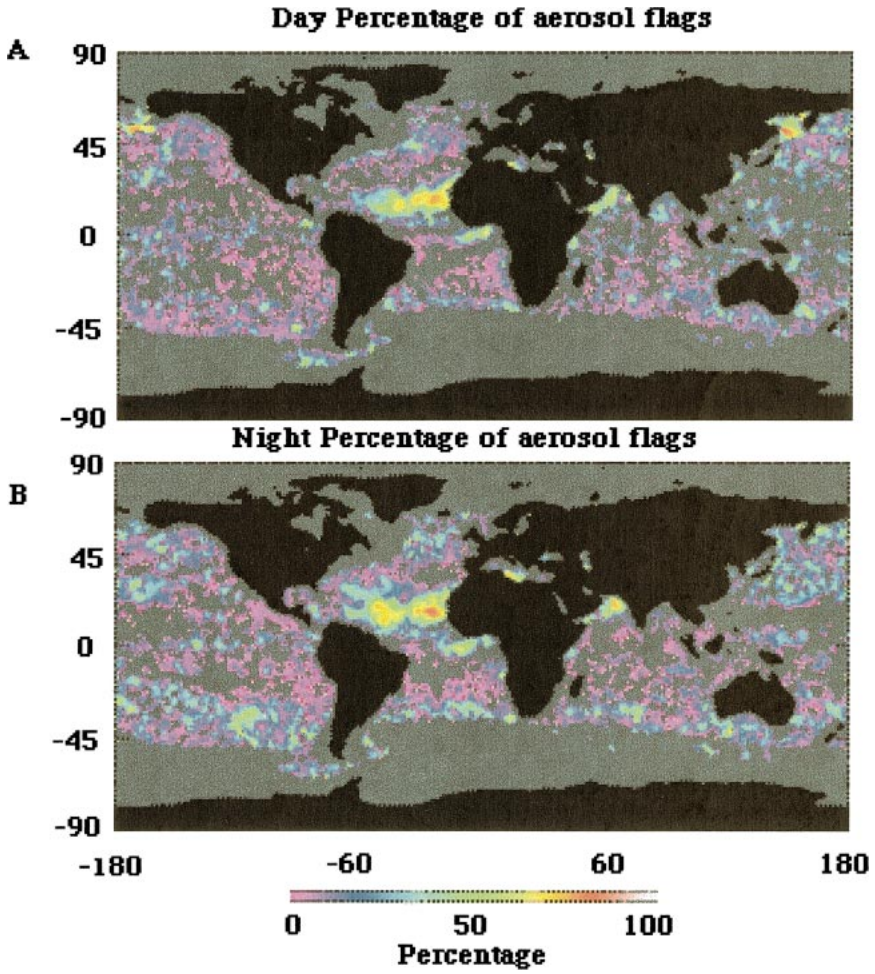

FIG. 6. (a) Spatial map showing percentage of pixel values flagged as aerosols for daytime MPFSST − ASST2 fields. Color scale ranges from 0 to 100 with purple corresponding to minimum values of 0 and red to white corresponding to 100. Gray values are points with no values where either there are too few collocated points or where the correlation was not statistically significant. (b) Same as (a) except for nighttime MPFSST − ASST2 values.

temperature difference (Murray et al. 1996; Donlon et al. 2002). Correlations of 0.20 certainly indicate that a large part of the variability can still be attributed to other sources. Results also show the importance of doing skin–bulk temperature comparisons on global scales, as regional differences due to atmospheric and oceanic conditions may be significant. Such comparisons are best performed using in situ measurements of skin temperatures with shipborne radiometers. To fully understand the remaining differences between the MPFSST and the ASST2 studies such as Murray et al. (1996) and Donlon et al. (2002) need to be expanded globally with in situ skin measurements. Donlon et al. (2002) showed that, in a limited area of the equatorial Pacific, under low speeds, a large bias developed with respect to both

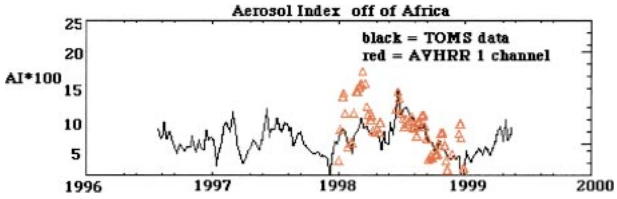

FIG. 7. Spatially averaged AI off the West African coast. The black line indicates AI from the TOMS data while the red triangles represent the AI from the 1-channel AVHRR algorithm.

TABLE 3. Differences between MPFSST, ASST2, and the WOD02 in situ data off the western coast of Africa.

Pass/comparison	Mean differences (°C)	Std dev (°C)	No. of points
MPFSST day flag 4	−0.36	0.65	625
MPFSST day flag 7	−0.28	0.64	373
MPFSST night flag 4	−0.48	0.63	367
MPFSST night flag 7	−0.33	0.55	196
ASST2 daytime	−0.20	0.46	234
ASST2 nighttime	−0.03	0.41	128

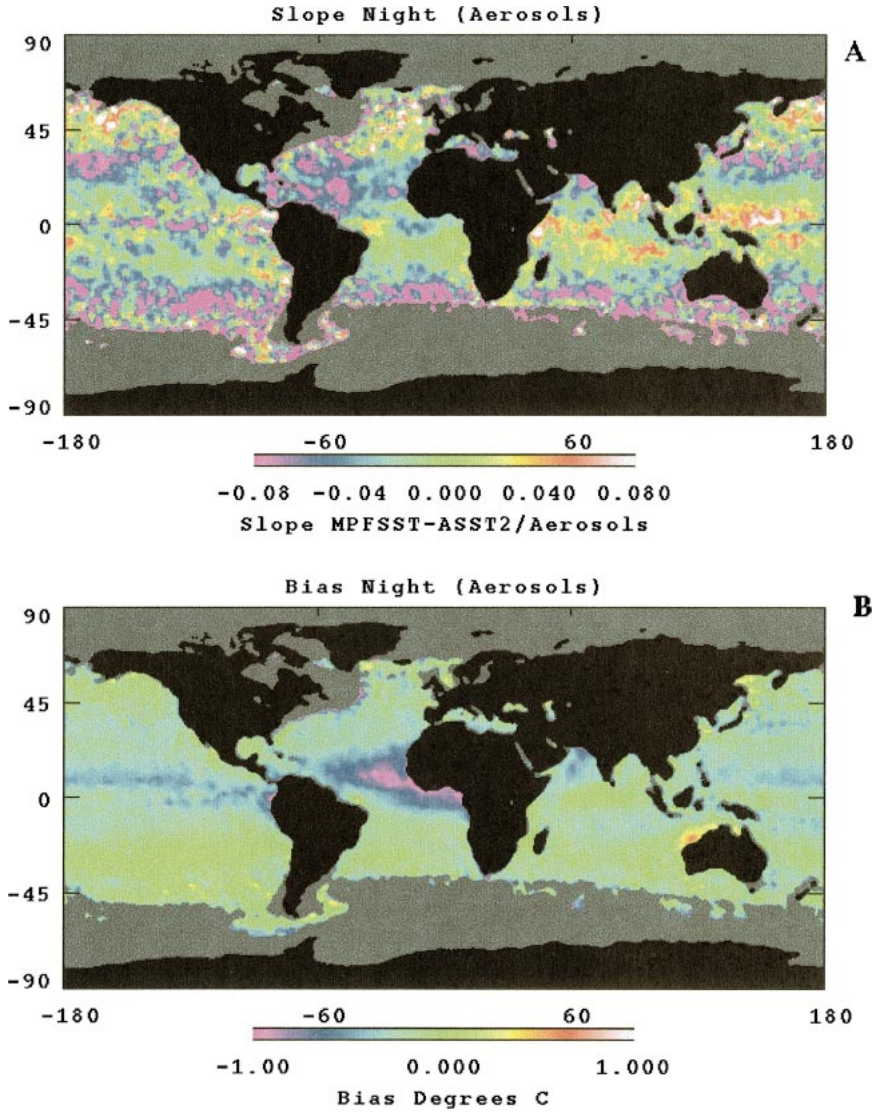

FIG. 8. (a) Slope calculated from the linear fit of MPFSST − ASST2 nighttime values aerosol. Color scale goes from −0.08 to 0.08 where minimum values are in purple and maximum values in red. (b) Bias calculated from the linear fit of MPFSST − ASST2 nighttime values aerosol. Color scale goes from minimum values of −1 to maximum values of 1.

the MPFSST and the ASST2 and SST measurements taken by the Tropical Atmosphere Ocean (TAO) array. At wind speeds greater than 2.5 m s^{-1} the MPFSST behaved more like a bulk temperature, while the ASST showed a cool skin temperature with respect to the depth measurements taken by the TAO array. Large error bars for comparisons done for low wind speeds indicate that remaining daytime differences of 0.10°C, with the ASST2 being cooler, would be consistent with the variability in the skin–bulk temperature differences seen during low wind speeds associated with Figs. 4a and 5a (see Donlon et al. 2002). Emery et al. (2001) have shown the importance of understanding the skin–bulk temperature differences for the implementation of any operational scheme for processing SST. Such differences must also be clearly understood before any merging strategy for the MPFSST and ASST can be implemented. This will necessitate using both global heat flux and wind data because of their influence on the skin–bulk temperature differences and the variability of such differences on global scales. Other sources of error such as water vapor, and the diurnal cycle (e.g., Gentemann et al. 2003) need to be examined.

9. Conclusions

First, it should be emphasized that the raw MPFSST − ASST2 results are a significant improvement over those obtained in a previous study (VCS) for MPFSST − ASST1. The reduction in the regional mean differ-

ences as defined by MPFSST − ASST2, after application of a simple linear flagging, is encouraging for the area off the western coast of Africa, where high aerosol loadings due to Saharan dust storms are known to exist. The nighttime bias and slope of the linear fit between the MPFSST − ASST2 and aerosols (see Fig. 8) indicate that in the region of high correlation off the western coast of Africa negative slopes and biases are consistent with increasing aerosols causing a cooler MPFSST versus ASST2. This is also supported by the WOD02 in situ comparisons as well.

Based on a cloud flagging procedure that uses data from the TOMS satellite, the area of the Arabian Sea and Bay of Bengal is shown to be still prone to possible residual cloud effects in either the MPFSST or the ASST2. Another possible explanation is that this is an area of upwelling that is not being picked up by the MPFSST because of the 2° constraint imposed by the cloud detection tests using the Reynolds optimal interpolation (OI) analysis. Further work needs to be done to differentiate between these possible scenarios in coastal areas. In general, though, correlations with the TOMS cloud data were insignificant, which is indicative of a significant improvement over the ASST1 results (see VCS). Comparisons between the MPFSST, the ASST2 and in situ data from the WOD02 in the area off the African coast demonstrate larger mean differences exist for the MPFSST. Nighttime mean differences for the ASST2 display the encouraging result of 0° ± 0.4°C, while the MPFSST is −0.3° ± 0.55°C.

Remaining differences should be examined with respect to skin–bulk temperature, possible diurnal effects, and water vapor. The coupling between the diurnal cycle and skin–bulk temperature differences needs to be appreciated, especially since the local time of overpass of AVHRR is different by some hours from that of ATSR. Such issues need to be addressed on a global basis, for example, by incorporation of heat flux and wind data into parameterizations that estimate their influence on skin–bulk temperature differences (e.g., Horrocks et al. 2003). A remaining mean difference of −0.2°C off the West African coast indicates improved aerosol flagging/correction schemes are required in order to achieve the 0.1°C accuracies needed for climate studies. Statistical models need to be examined for the possibility of applying the TOMS aerosol data in correction algorithms, as well as flagging. These tests should address the role of quality flags in climate data records. The difficulty with flagging is that data are simply excluded and, as tropospheric dust and smoke aerosol exhibit seasonal cycles, this may alias the SST signal. Thus, correction schemes, such as those recently developed for operational AVHRR SST retrievals (e.g., Nalli and Stowe 2002), are preferable and should be adapted and applied retrospectively. Such correction schemes are being examined as reprocessing efforts of the MPFSST time series are examined. Additionally, such schemes, because they incorporate an AVHRR versus TOMS derived aerosol optical depth avoid possible errors due to the collocation of the two sensors. It is important to remember that, although significant, correlations of 0.3 indicate that greater than 50% of the variability of the MPFSST − ASST2 is still not being explained by aerosols or clouds. Thus, future work needs to focus on areas where the skin–bulk temperature differences between the MPFSST and the ASST2 are significant. A clear understanding of these issues are critical before the datasets can be used for climate studies.

Nonetheless, the improvement in the comparisons of the MPFSST and the ASST2 after the application of the simple flagging for aerosols and clouds indicates continued thought and analysis must be given to applications of cloud detection and/or aerosol removal scheme. Additionally, the viability of using additional remote sensing data sets in the quality control of satellite-based SST retrievals is demonstrated.

Acknowledgments. This work was carried out under contract with the National Aeronautics and Space Administration at the Jet Propulsion Laboratory, California Institute of Technology. Jay Herman at the Goddard Space Flight Center is kindly thanked for the original idea of comparing the Pathfinder SST data and the ASST2 with the TOMS aerosol and cloud data. Christopher Mutlow at the Rutherford Appleton Laboratory is thanked for kindly providing access to the ATSR2 data. We would also like to thank an anonymous reviewer for a thoughtful critique of this paper.

REFERENCES

Allen, M. R., C. T. Mutlow, G. M. C. Blumberg, J. R. Christy, R. T. McNider, and D. T. Llewellyn-Jones, 1994: Global change detection. *Nature*, **370,** 24–25.

Chiapello, I., P. Goloub, D. Tanre, A. Marchand, J. Herman, and O. Torres, 2000: Aerosol detection by TOMS and POLDER over oceanic regions. *J. Geophys. Res.*, **105,** 7133–7142.

Conkright, M. E., and Coauthors, 2002: *Introduction.* Vol. 1, *World Ocean Database 2001,* NOAA Atlas NESDIS 42, 167 pp.

Donlon, C., 2002: The GODAE High Resolution Sea Surface Temperature Project. *Proc. GODAE Data and Product Server Workshop*, Biarritz, Fance, GODAE, 16 pp.

——, P. J. Minnett, C. Gentemann, T. J. Nightingale, I. J. Barton, B. Ward, and M. J. Murray, 2002: Toward improved validation of satellite sea surface skin temperature measurements for climate research. *J. Climate*, **15,** 353–369.

Emery, W. J., S. Castro, G. A. Wick, P. Schluessel, and C. Donlon, 2001: Estimating sea surface temperature from infrared satellite and in situ temperature data. *Bull. Amer. Meteor. Soc.,* **82,** 2773–2785.

Gentemann, C. L., C. J. Donlon, A. Stuart-Menteth, and F. J. Wentz, 2003: Diurnal signals in satellite sea surface temperature measurements. *Geophys. Res. Lett.,* **30,** 1140, doi:10.1029/2002GL016291.

Horrocks, L. A., B. Candy, T. J. Nightingale, R. W. Saunders, A. O'Carroll, and A. R. Harris, 2003: Parameterizations of the ocean skin effect and implications for satellite-based measurement of sea-surface temperature. *J. Geophys. Res.,* **108,** 3096, doi: 10.1029/200.

Jones, M. S., and M. A. Saunders, and T. H. Guymer, 1996: Reducing

cloud contamination in ATSR averaged sea surface temperature data. *J. Atmos. Oceanic Technol.,* **13,** 492–506.

Kilpatrick, K. A, G. P. Podesta, and R. Evans, 2001: Overview of the NOAA/NASA Pathfinder algorithm for sea surface temperature and associated matchup database. *J. Geophys. Res.,* **106,** 9179–9197.

May, D. A., L. L. Stowe, J. D. Hawkins, and E. P. McClain, 1992: A correction for Saharan dust effects on satellite sea surface temperature measurements. *J. Geophys. Res.,* **97,** 3611–3619.

McPeters, R. D., P. K. Bhartia, A. J. Krueger, and J. R. Herman, 1998: Earth Probe Total Ozone Mapping Spectrometer (TOMS) data products user's guide. NASA Tech. Publication 206895, GSFC, Greenbelt, MD, 23 pp.

Merchant, C. J., and A. R. Harris, 1999: Toward the elimination of bias in satellite retrievals of sea surface temperature 2. Comparison with in situ measurements. *J. Geophys. Res.,* **104C,** 23 579–23 590.

——, ——, M. J. Murray, and A. M. Závody, 1999: Toward the elimination of bias in satellite retrievals of sea surface temperature 1. Theory, modeling and interalgorithm comparison. *J. Geophys. Res.,* **104,** 23 565–23 578.

Murray, M. J., M. R. Allen, and C. T. Mutlow, 1996: Global evaluation of sea surface temperature retrievals from the along-track scanning radiometer through comparison with the NOAA operational analysis. Tech. Rep. RAL-TR-96-60, 12 pp.

Nalli, N. R., and L. L. Stowe, 2002: Aerosol correction for remotely sensed sea surface temperatures from the National Oceanic and Atmospheric Administration Advanced Very High Resolution Radiometer. *J. Geophys. Res.,* **107,** 3172, doi:10.1029/2001JC001162.

Reynolds, R. W., and T. M. Smith, 1994: Improved global sea surface temperature analysis using optimum interpolation. *J. Climate,* **7,** 929–948.

Simpson, J. J., T. J. McIntire, J. R. Stitt, and G. L. Hufford, 2001: Improved cloud detection in AVHRR daytime and night-time scenes over the ocean. *Int. J. Remote Sens.,* **22,** 2585–2615.

Vazquez-Cuervo, J., and R. Sumagaysay, 2001: A comparison between sea surface temperatures as derived from the European Remote Sensing Along-Track Scanning Radiometer and the NOAA/NASA AVHRR Oceans Pathfinder dataset. *Bull. Amer. Meteor. Soc.,* **82,** 925–944.

Walton, C. C., 1988: Nonlinear multichannel algorithms for estimating sea surface temperature with AVHRR satellite data. *J. Appl. Meteor.,* **27,** 115–124.

Závody, A. M., C. T. Mutlow, and D. T. Llewellyn-Jones, 1995: A radiative transfer model for sea surface temperature retrieval for the along-track scanning radiometer. *J. Geophys. Res.,* **100,** 937–952.