Satellite-Derived Sea Surface Temperatures: Evaluation of *GOES-8* and *GOES-9* Multispectral Imager Retrieval Accuracy

DOUGLAS A. MAY AND WALTER O. OSTERMAN

Naval Research Laboratory, Remote Sensing Applications Branch, Stennis Space Center, Mississippi

(Manuscript received 27 January 1997, in final form 28 July 1997)

ABSTRACT

Sea surface temperature (SST) retrieval accuracy from the multispectral imager on the new generation of GOES satellites is analyzed. Equations for two and three infrared channels are empirically derived using cloud-free satellite radiances matched to buoy SST measurements obtained in 1995 and 1996. Both *GOES-8* and *GOES-9* demonstrate the capability to retrieve sea surface temperature at better than 1-K root-mean-square difference (rmsd) with negligible bias relative to buoy SST measurements. *GOES-8* rmsd errors are found to be 0.79 K (day) and 0.81 K (night). *GOES-9* rmsd errors are 0.65 K (day) and 0.59 K (night). The *GOES-9* results are relatively comparable to those currently achieved operationally from the NOAA polar-orbiting satellite Advanced Very High Resolution Radiometer sensor. Investigation revealed that GOES imager multiple detector scan striping impacted SST accuracy, requiring sample array averaging for best results.

1. Introduction

Satellite sea surface temperature (SST) retrieval has been studied extensively (Anding and Kauth 1970; Maul and Sidran 1972; McMillin 1975; McMillin and Crosby 1984; Barton et al. 1989; Barton 1995). Utilization of multiple channel infrared (IR) satellite sensors has demonstrated the most success. McMillin and Crosby (1984) provide an overview of the fundamental approach. McClain et al. (1985) describe in detail the original methods employed for operational retrieval of SST from the polar-orbiting National Oceanic and Atmospheric Administration (NOAA) Advanced Very High Resolution Radiometer (AVHRR). Furthermore, Barton (1995) provides an extensive review of more recent satellite SST retrieval methods and developments. Global root-mean-square-difference (rmsd) errors of 0.6–0.7 K (Strong and McClain 1984; McClain 1989; May et al. 1992) can be realized if the data is properly cloud screened and void of high aerosol dust concentrations. This accuracy enables operational use within oceanographic analyses, atmospheric forecasts, climate studies, and civilian and military maritime activities (Clancy et al. 1992; Reynolds and Smith 1994; Hogan and Rosmond 1991; Smith et al. 1996; Barnston and Ropelewski 1992).

Early attempts at retrieving sea surface temperature from Geostationary Operational Environmental Satellite (GOES) sensors were significantly limited by the single IR channel, spatial resolution, and channel digitization levels available (Maul et al. 1978; Maul 1981). These limitations hindered the ability to remove atmospheric effects from the received emissions and restricted the thermal sensitivity of the IR channel. Bates and Smith (1985) demonstrated some success with the *GOES-5* Visible and Infrared Spin-Scan Radiometer (VISSR) Atmospheric Sounder (VAS), combining the 3.9-μm sounding channel with the 11-μm channel. SST retrieval errors of 0.9–1.0 K were attained. However, IR channel spatial resolution was 8–16 km, and sensor acquisition mode schedules limited timely access to some regions. Consequently, SST retrieval from a satellite has primarily focused on utilizing the higher resolution (1–4 km) 10-bit multichannel IR data available from polar-orbiting satellite sensors. This study evaluates the potential to accurately retrieve SST from the latest generation of GOES satellites.

2. Potential *GOES-8* and *GOES-9* SST retrieval accuracy and application

GOES-8 (75°W) and *GOES-9* (135°W) are equipped with the latest generation geostationary imaging sensor (Menzel and Purdom 1994). This sensor provides a significant improvement to the IR scanning channels on previous GOES satellites. The three IR channels now present on the new GOES imager are located in IR atmospheric water vapor ''windows'': channel 2 (3.7–3.9 μm), channel 4 (10.2–11.2 μm), and channel 5 (11.5–12.5 μm). These three channels provide 10-bit multichannel IR data at 4-km spatial resolution and ap-

Corresponding author address: Dr. Douglas A. May, NRL Code 7240, Remote Sensing Applications Branch, Naval Research Laboratory, Stennis Space Center, MS 39529-5004.
E-mail: may@snaps.nrlssc.navy.mil

TABLE 1. Infrared channel NEΔT (K) for NOAA and GOES imagers.

Satellite	3.7 μm	11 μm	12 μm
NOAA-14	0.25	0.035	0.05
GOES-8	0.17	0.12	0.21
GOES-9	0.13	0.07	0.155

proximately 0.12-K thermal sensitivity, comparable to AVHRR IR channels and subsequently well suited for SST retrieval.

In addition to digital thermal sensitivity, sensor channel noise level is also important to satellite SST retrieval accuracy. AVHRR SST retrieval accuracy can in part be attributed to the IR channel thermal sensitivity plus the relatively low channel noise equivalent temperature difference (NEΔT). AVHRR IR channel NEΔT specifications are given by NOAA as ± 0.12 K at 300 K (Schwalb 1978, 1982). However, AVHRR 11- and 12-μm channels have consistently demonstrated noise levels better than ± 0.05 K (Dudhia 1989). Table 1 depicts the IR channel NEΔT noise levels for *NOAA-14* AVHRR taken from NESDIS along with the three IR channels of *GOES-8* and *GOES-9* (M. Weinreb 1996, personal communication). Although the GOES 11- and 12-μm channel NEΔT levels are somewhat larger than those experienced for AVHRR, we will show that they are sufficient to retrieve SST at better than 1-K rmsd accuracy. Also note that the *GOES-8* and *GOES-9* 3.7-μm channel NEΔT levels are lower than those for *NOAA-14*. Relatively high 3.7-μm channel noise has been a problem for all AVHRR sensors with the exception of *NOAA-11* (May et al. 1993).

Equation (1) represents the daytime split-window (uses only the 11- and 12-μm IR channels) multichannel sea surface temperature (MCSST) algorithm for *NOAA-14* satellites presently operational at the Naval Oceanographic Office (NAVOCEANO):

$$\text{MCSWD} = 1.0135T_{11} + 2.2014(T_{11} - T_{12}) + 0.7833(T_{11} - T_{12})(\sec\theta - 1) - 277.4234, \quad (1)$$

where MCSWD represents the multiple channel split-window daytime SST (°C), T_{11} represents the 11-μm brightness temperature (K), T_{12} represents the 12-μm channel brightness temperature (K), and $\sec\theta$ is the secant of the satellite zenith angle.

NOAA-14, GOES-8, and *GOES-9* 11- and 12-μm channel NEΔT levels from Table 1 can be applied to Eq. (1) to estimate the magnitude of temperature retrieval uncertainty contributed by channel noise. These results are presented in Table 2. Calculations were performed assuming a surface SST of 300 K at nadir view. Results for the operational nighttime algorithm were also calculated and are displayed in Table 2. The operational NAVOCEANO nighttime *NOAA-14* algorithm utilizes the 3.7-μm channel and is presented as

TABLE 2. Estimated satellite SST retrieval error attributed to channel noise for daytime (MCSWD) and nighttime (MCTWN) algorithms.

Satellite	MCSWD	MCTWN
NOAA-14	0.22	0.31
GOES-8	0.85	0.47
GOES-9	0.57	0.33

$$\text{MCTWN} = 1.0082T_{11} + 0.9167(T_3 - T_{12}) + 1.7044[\sec\theta - 1] - 274.8001, \quad (2)$$

where MCTWN is the multiple channel triple-window (uses the three IR channels 3.7, 11, and 12 μm) nighttime SST (°C) and T_3 represents the 3.7-μm channel (K).

Table 2 reveals that IR channel noise contributes approximately 0.2–0.3-K uncertainty (0.22 K daytime and 0.31 K nighttime) to the *NOAA-14* AVHRR sensor MCSST retrieval estimate. This uncertainty is significantly higher for *GOES-8* and *GOES-9* due to increased NEΔT levels in their sensor channels. However, except for the *GOES-8* split-window daytime algorithm, all uncertainties are less than 0.6 K.

Figure 1 displays the total rmsd error with respect to drifting buoy SST measurements associated with operational NAVOCEANO *NOAA-14* daytime and nighttime algorithms from April 1995 (the first full month of operational *NOAA-14* SST data) through September 1996. The average daytime rmsd for this time period is calculated to be 0.54 K. The average nighttime rmsd for this same time period is 0.50 K. From Table 2, approximately 0.22 K of the total daytime rmsd can be attributed to sensor channel NEΔT levels. The remaining error is primarily attributed to atmospheric correction uncertainties, skin/bulk temperature differences, and subpixel cloud contamination (Harris and Saunders 1996). Satellite–buoy matchup time and distance dif-

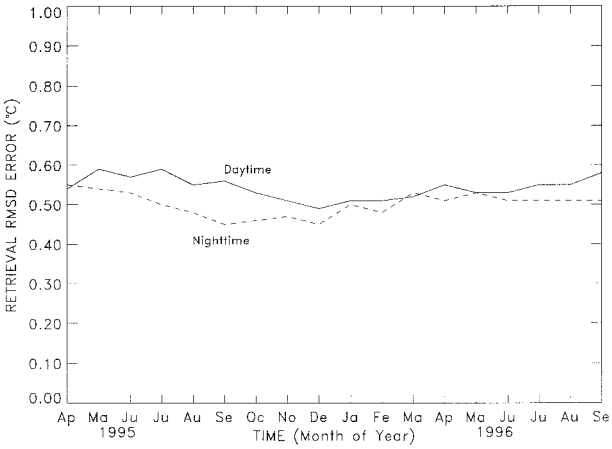

FIG. 1. Root-mean-square difference error for *NOAA-14* daytime and nighttime retrievals relative to global drifting buoy data between April 1995 and September 1996.

TABLE 3. Estimated satellite SST total error for daytime (MCSWD) and nighttime (MCTWN) algorithms.

Satellite	MCSWD	MCTWN
NOAA-14	0.54	0.50
GOES-8	0.98	0.61
GOES-9	0.74	0.52

ferences also provide a source of statistical error (Minnett 1991). The remaining rmsd error can be quantified by the following formula:

$$E_{\text{rem}} = (E_{\text{tot}}^2 - E_{\text{ch}}^2)^{1/2}, \quad (3)$$

where E_{rem} represents the remaining error, E_{tot} is the total error (Fig. 1), and E_{ch} represents the IR channel noise error (Table 2). The remaining daytime rmsd error is then calculated to be 0.49 K. For nighttime algorithms, the channel noise error is 0.31 K and the remaining rmsd error calculated to be 0.39 K.

Since AVHRR and GOES imager channel bandwidths are relatively similar, we assume that the preceding principles are applicable for estimating GOES SST retrieval error. Adding the AVHRR SST remaining rmsd error values (E_{rem}) to the *GOES-8* and *GOES-9* channel noise demonstrates that the imager on *GOES-9* should be capable of retrieving a daytime split-window SST at an accuracy of about 0.74 K rmsd (Table 3). The *GOES-8* split-window accuracy estimate is slightly worse at 0.98 K rmsd. The nighttime triple window accuracies are significantly better for both satellites: 0.61 K for *GOES-8* and 0.52 K for *GOES-9.* The latter values are encouraging since they approach the rmsd range typically associated with AVHRR data. The following sections present results from real GOES satellite data matched to buoy SST measurements that confirm these theoretical accuracies.

GOES-8 and *GOES-9* therefore provide the potential of retrieving SST at better than 1-K thermal accuracy over large regions in a very timely manner (every half hour around the continental United States and North American sectors and every 3 h for full disk hemispheric scenes). This refresh rate far surpasses that presently available from the polar orbiters. Multiple images also provide the opportunity to effectively cloud screen large regions by temporal compositing. The probability of obtaining a cloud-free view of the ocean surface in a specified area is, therefore, greatly enhanced. The capability to better quantify temporal SST changes, such as diurnal warming events over regional and ocean basin areas, is also a possibility from geostationary data. Such potential has not been fully realized with polar-orbiting sensor data. More timely SST information will further our understanding of air–ocean flux exchange processes and improve atmospheric and oceanographic model accuracy, particularly tropical cyclone and coupled air–ocean models to which these thermodynamic processes are so important (Emmanuel 1988; Gallacher et al. 1989).

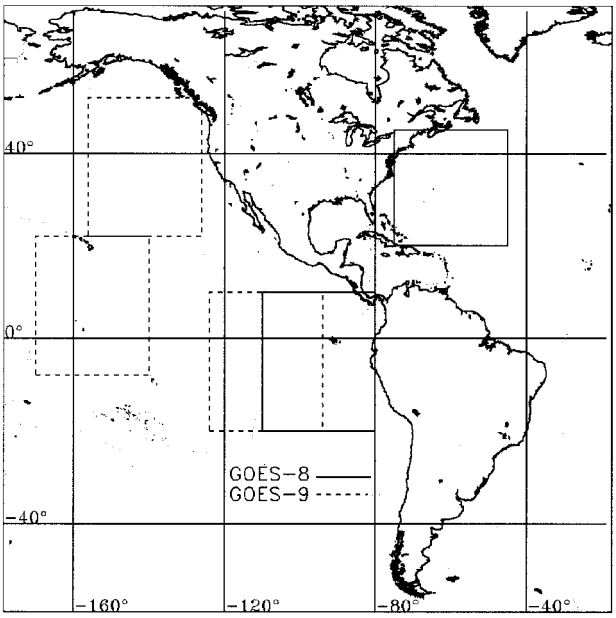

FIG. 2. Geographic location of the *GOES-8* and *GOES-9* data sectors used in the SST accuracy analysis.

3. Data collection and processing

The *GOES-8* and *GOES-9* data used in this study were obtained through the NRL-Monterey Satellite Meteorological Applications Section satellite data receiving system. Digital full disk hemispheric datasets from each satellite were utilized to provide a diverse sample of surface temperatures, atmospheric conditions, and satellite scan angles. Preselected regional sectors were extracted from the digital data received at NRL-Monterey. These data sectors were then automatically transferred via ftp to NRL-Stennis in near real time for archive and subsequent analysis. The sector regions for each satellite were carefully chosen to cover tropical and mid- to high-latitude areas (Fig. 2).

The digital data files contain calibrated and earth-located IR channel brightness temperatures at full 4-km spatial resolution and visible channel reflectances averaged to 4 km for data consistency. In addition, separate files containing coincident latitude, longitude, satellite zenith angle, solar zenith angle, and relative azimuth angle for each pixel location were also generated and archived. The latter two files were used for day/night determination and cloud detection purposes.

The *GOES-8* data sectors covered two separate geographic areas: one from the midlatitude western Atlantic and one from the tropical eastern Pacific (Fig. 2). Sector coordinates were selected based on optimum access to available moored and drifting buoy locations, while also diversely sampling surface temperatures, atmospheric conditions, and satellite scan angles. The western Atlantic area extended from 20° to 45°N and from 75° to 45°W. For this box, the satellite zenith angles varied between 23° and 60°. Buoy surface temperatures ranged

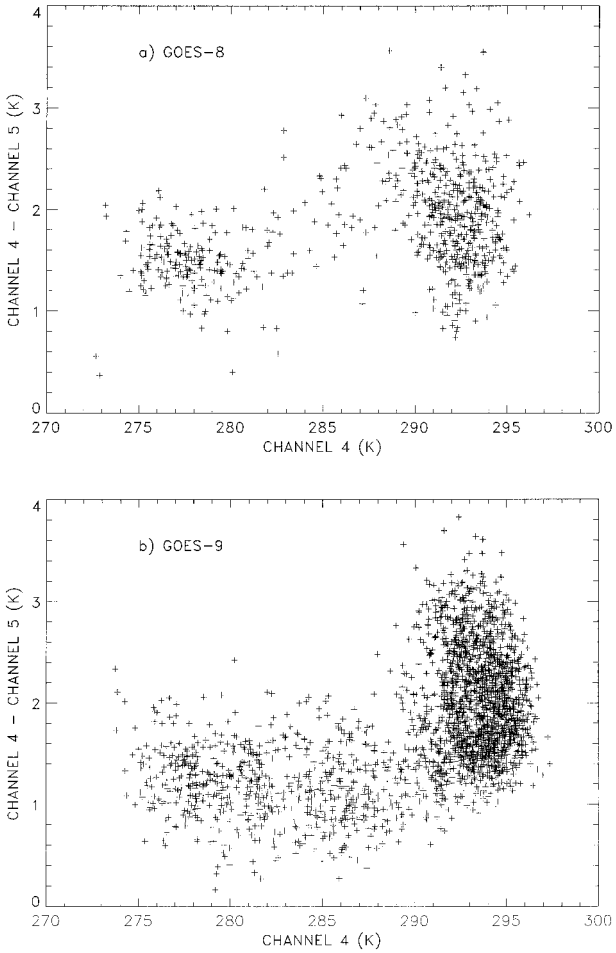

FIG. 3. (a) *GOES-8* channel 4 brightness temperature versus channel 4 minus channel 5 brightness temperature for all satellite–buoy matches; (b) same as in (a) except for *GOES-9*.

from less than 2° to 27°C. Images were collected in this box beginning 21 November 1995 and ending 3 January 1996.

The *GOES-8* tropical Pacific area extended from 20°S to 10°N and from 110° to 80°W. For this box, the satellite zenith angles covered from near nadir to 45°. Buoy surface temperatures were typically warm, varying from 20° to 29°C. Images were obtained for this box beginning 6 December 1995 and ending 3 January 1996.

Figure 3a depicts the range of all satellite–buoy matchup *GOES-8* channel 4 brightness temperatures versus the channel 4 minus 5 difference. The latter provides an estimate of atmospheric water vapor content encountered in the samples. The figure demonstrates that a diverse range of surface temperatures and atmospheric water vapor conditions were collected.

The *GOES-9* data covered three geographic areas, all located in the Pacific Ocean. These regions were also carefully selected to achieve a wide range of surface temperatures, scan angles, and SST measurements. These boxes are designated as North Pacific, central Pacific, and eastern Pacific. The North Pacific box reached from 22° to 52°N and from 156° to 126°W. For this area, the satellite zenith angles varied between 25° and 66°. Buoy surface temperatures ranged from 5° to 23°C. Images were collected for this area beginning 10 January 1996 and ending 8 April 1996.

The central Pacific box extended from 8°S to 22°N and from 170° to 140°W. For this area, the satellite zenith angles covered from near nadir to 47°. Buoy surface temperatures varied between 22° and 30°C. Images were obtained in this area beginning 27 February 1996 and ending 8 April 1996.

The eastern Pacific box reached from 20°S to 10°N and from 124° to 94°W. For this area, the satellite zenith angles varied between 13° and 52°. Buoy surface temperatures ranged from 20° to 29°C. Images were collected in this area beginning 10 January 1996 and ending 26 February 1996.

Figure 3b shows the range of all satellite–buoy matchup *GOES-9* channel 4 brightness temperatures versus the channel 4 minus 5 difference. As with *GOES-8*, the figure demonstrates that a wide range of surface temperatures and atmospheric water vapor conditions were collected.

Global drifting and moored buoy measurements were obtained from NAVOCEANO where these observations are routinely collected as part of the operational AVHRR satellite SST quality control processing. Buoy SSTs were matched to cloud screened GOES satellite data by implementing a two-pass process. The cloud detection process is described in subsequent paragraphs below. During the first pass, the satellite images were indexed according to year, month, and day to facilitate processing. Buoy measurement times from the buoy files were then identified and matched to the nearest corresponding image in time. Buoy locations were used to identify corresponding pixel locations within each image. The matched buoy measurements and image pixel values were then saved in a match file. In this first processing pass, all the buoy data were matched to the image data such that the buoy SST and the satellite SST were separated in time by no more than 4 h and in space by no more than one pixel.

Each match was then checked for image pixel cloud contamination. The match file was sorted according to satellite image so that each image was cloud screened only once. Within this second processing pass, only those image pixels in the local vicinity of the buoy match location were checked for cloud contamination. This minimized image processing run time for each image. Cloud-free image pixels located within approximately 25 km of buoy match locations were considered valid satellite–buoy matches. However, only the closest cloud-free pixels were retained for each individual buoy match. The surviving cloud-free matches, combined with all pertinent buoy measurements, times, and locations as well as image pixel brightness temperatures, times, and locations, were the basis of the SST analysis.

Combinations of visible and IR channels were utilized to determine whether image pixels were cloud contaminated. The cloud detection tests were separated into day and night categories that essentially follow the concepts described by McClain et al. (1985). Some minor differences and threshold changes were necessary to conform to the GOES IR scanner channels available.

The *GOES-8* and *GOES-9* IR imager channels consist of two detectors oriented in a north–south array that scans in oscillation west–east and then east–west (Komadja and McKenzie 1994). Alternating lines in the image data are observed by the two detectors, generating some scan striping in the imagery caused by detector output differences. Each detector is independently calibrated and striping of several tenths of a kelvin have been observed (Baucom and Weinreb 1996). Several pixel array averages centered on the buoy-image pixel match location were calculated to investigate how well pixel averaging can minimize imager scan striping while maximizing local satellite temperature accuracy. The pixel averaging combinations tested included the single pixel, 3×3, 5×5, 7×7, 9×9, 11×11, 13×13, and 15×15 pixel array averages. Averaging over an odd number of lines and pixels was performed to maintain geolocation of a central pixel for direct comparison to single-pixel processing accuracy. Only cloud-free pixels were averaged. With the GOES imager IR channel field of view at 4 km, pixel array averages covering approximately 4–60 km were evaluated. It will be demonstrated in the following sections that a 9×9 pixel average provides the most accurate results. Utilization to higher resolutions down to 3×3 averages contributed no more than 0.05-K additional error. Investigation of a processing filter to remove scan striping should be pursued. Such efforts could increase imager thermal accuracy at high resolution.

4. Comparison method

The buoy and satellite matches were used to predict the buoy SST measurement from the satellite channel brightness temperatures by regressing the standard SST equation form,

$$\mathrm{SST} = aT_i + b(T_i - T_j) + c(\sec\theta - 1) + d, \quad (4)$$

to the buoy SST measurements. Here, T_i and T_j represent GOES channel brightness temperatures measured in channels i and j, and a, b, c, and d are coefficients determined by regression.

GOES-8 and *GOES-9* data were kept separate in order to derive individual equations for each satellite. In addition, each satellite's daytime and nighttime data were also kept separate for cloud screening purposes and to investigate channel 2 utilization within nighttime SST algorithms. The day and night data were also combined to generate a 24-h algorithm as well.

The final data matches were divided into training and test sets to derive and evaluate empirical MCSST algorithms. The training set consisted of half the matchups and was selected from every other match within the file. The MCSST equations were derived from this training set through regression procedures. The test set consisted of the other half of the matches. This dataset was used to test the accuracy of the regression relationship on totally independent data. Equations were derived and evaluated for various pixel array averages ranging from single pixels to 15×15 sample averages. The accuracies of complete 24-h, daytime only, and then nighttime only algorithms were compared and evaluated.

TABLE 4. *GOES-8* SST retrieval regression training and test statistics.

Pixels averaged	Multiple cross correlation	Standard error of estimate (K)	Adjusted R^2
Training set (529 matches)			
1×1	0.9969	0.88	0.9938
3×3	0.9971	0.85	0.9942
5×5	0.9972	0.84	0.9944
7×7	0.9973	0.82	0.9947
9×9	0.9974	0.80	0.9949
11×11	0.9973	0.81	0.9947
13×13	0.9966	0.92	0.9933
15×15	0.9964	0.94	0.9929
Pixels averaged		**Bias (K)**	**rmsd (K)**
Test set (545 matches)			
1×1		0.04	0.91
3×3		0.01	0.87
5×5		0.02	0.86
7×7		0.02	0.85
9×9		0.04	0.84
11×11		0.04	0.87
13×13		0.05	0.91
15×15		0.06	0.97
Equation (5) for various pixel array averages (545 matches)			
1×1		−0.08	0.93
3×3		−0.07	0.87
5×5		−0.03	0.86
7×7		−0.00	0.85
9×9		0.04	0.84
11×11		0.06	0.87
13×13		0.07	0.91
15×15		0.07	0.97

5. *GOES-8* SST algorithm results

a. 24 hours

Regression of the split-window MCSST equation to the entire 24-h dataset for *GOES-8* indicates that the best fit occurs when using a 9×9 pixel average (Table 4). Table 4 shows that 99.49% of the variation in SST (R^2) is explained by this model. In addition, the standard error of estimate relative to the training set is 0.80 K. These statistics are better than those obtained by other sample array averages. The other models do relatively well though with the next best fit occurring with an 11

$\times$ 11 pixel average. This model exhibits a standard error of 0.81 K. A 7 $\times$ 7 pixel average registers the third best fit with a standard error of 0.82 K.

The regression fit worsens as fewer image pixels are used in the sample averaging process. The regression model that utilizes only the single center pixel shows a standard error of 0.88 K, almost 0.1 K worse than the 9 $\times$ 9 regression algorithm results. This increased scatter is most likely a result of scan striping in the data caused by the multiple IR detector array. Using sample array averages greater than 9 also worsens the regression fit obtained. This consequence is thought to be due to inhomogeneous surface temperature signals or natural gradients across large ocean distances (greater than 44 km). Additional pixel averaging therefore reduces the accuracy of these retrieval estimates relative to the true temperature at the buoy location.

Averaging 9 $\times$ 9 pixels appears to perform best for removing the striping and scan noise while also retaining a good estimate of surface temperature at the buoy. A consequence of a 9 $\times$ 9 sample average is that it will impact the user's ability to discern small-scale ocean features of potential interest. The best-fit 24-h regression algorithm for *GOES-8* is presented below:

$$\text{G8SST} = 1.075T_{11} + 1.8247(T_{11} - T_{12}) + 2.1698(\sec\theta - 1) - 294.6801, \quad (5)$$

where G8SST represents the *GOES-8* 24-h split-window SST (°C).

Tests of all the various array size regression equations to the appropriate test dataset demonstrate that the standard deviations are somewhat higher that those obtained for the training set (Table 4). This outcome is to be expected since the test dataset has not been used to train the regression equations. Statistics show that the 9 $\times$ 9 pixel average equation performs the best with an overall rmsd of 0.85 K. Again, rmsd statistics decrease in accuracy as fewer pixels are averaged into the process or if more pixels are used. Bias statistics are remarkably less than 0.06 K for all pixel-averaged algorithms.

Since users may wish to process GOES data using various sample array sizes, the accuracy of the 9 $\times$ 9 equation on other pixel array averages was investigated. The results are presented in Table 4. The rmsd calculations are only slightly worse those obtained previously in the table and in some instances better. The bias errors show that the 9 $\times$ 9 pixel regression equation is slightly warm for sample arrays involving more than 9 pixels and slightly cold for less sample averaged arrays. The results in Table 4 indicate that the 9 $\times$ 9 regression equation is the most accurate 24-h equation and that it can also be used for various sample array averages without significant addition of error. Processing 5 $\times$ 5 or 3 $\times$ 3 pixel-averaged data will only add about 0.02–0.03 K to the SST error. This additional error is relatively small, thus processing at this resolution will still provide decent accuracy and more oceanographic detail. Single-pixel SST processing is not recommended if accurate results are desired.

TABLE 5. *GOES-8* SST retrieval algorithm standard error (K).

Algorithm	Training set	Test set
G8SST	0.80	0.84
MCSST without $\sec\theta - 1$	0.82	0.87
MCSST with (T4–T5) $(\sec\theta - 1)$	0.81	0.86
NLSST	0.81	0.86

Alternative algorithm forms were investigated and compared to Eq. (5) results. Algorithms tested included one that neglects the satellite zenith angle term, one that takes the form of Eq. (1) [using a $(T_{11} - T_{12})(\sec\theta - 1)$ term], and also the nonlinear sea surface temperature (NLSST) algorithm described by Walton et al. (1990). Each of these algorithm results, along with the Eq. (5) results, appear in Table 5.

Table 5 shows that the regression standard errors are comparable, with G8SST providing the best results. A split-window equation that neglects the satellite zenith angle performs slightly worse than one that includes this term. Statistics from such an algorithm exhibit a standard error of 0.82 K on the training set and 0.87 K on the test set. This outcome is comparable to but slightly worse than the algorithm results obtained by using the satellite zenith angle term. To test whether the satellite zenith angle term is statistically significant, the critical *t* statistic is calculated to be 1.645 for 527 degrees of freedom at the 95% confidence level. The regression model satellite zenith angle term coefficient *t* statistic is found to be 4.068, which rejects the hypothesis that the true coefficient is equal to zero. Including the term within the regression is therefore found to be statistically significant.

From Table 5, regression of a split-window equation following the form of Eq. (1) performs slightly better than the equation that ignores the satellite zenith angle altogether. However, this equation still performs slightly worse than Eq. (5). The regression model *t* statistic for the coefficient of this term is found to be 2.408. Since the critical *t* statistic is 1.645, this term is statistically significant within the regression but not as significant as the term using the satellite zenith angle term alone, as presented in the preceding paragraph. Both algorithms are quite good; however, G8SST is more statistically significant. The training set shows a standard error of 0.81 and 0.86 K for the test set. The differences between these results and those for G8SST may be due to additional channel noise caused by the introduction of $(T_{11} - T_{12})$ within the term.

AVHRR data have demonstrated that use of a surface temperature estimate in the SST retrieval equation actually improves the retrieval accuracy (Walton et al. 1990). This NLSST technique utilizes a surface temperature value within the *b* parameter of Eq. (4) to nonlinearly correct atmospheric effects according to surface temperature variations. In this study, the surface tem-

TABLE 6. *GOES-8* daytime and nighttime algorithm statistics.

Dataset	Bias (K)	rmsd (K)
Daytime split		
Training (271 matches)	0	0.74
Test (277 matches)	−0.01	0.79
Nighttime split		
Training (258 matches)	0	0.84
Test (268 matches)	0.08	0.90
Nighttime triple		
Training (258 matches)	0	0.75
Test (268 matches)	0.01	0.81
Nighttime dual		
Training (258 matches)	0	0.82
Test (268 matches)	0.01	0.87

perature value used for testing an NLSST approach was obtained by using the resultant MCSST value from Eq. (5).

For *GOES-8*, the NLSST regression demonstrates no improvement relative to the MCSST equation (Table 5). The training set shows a standard error of 0.81 K with the test set error at 0.86 K. Rather than improving upon the MCSST accuracy, the NLSST statistics are actually about 0.01 K worse. For the data analyzed in this study, the GOES imager retrieval algorithm does not exhibit an atmospheric correction dependence on surface temperature similar to that of the AVHRR sensor data.

The *GOES-8* NLSST regression model *t* statistic for the surface temperature term was calculated to be 18.119, which is statistically significant at the 95% confidence level. However, the MCSST regression model *t* statistic for the same term excluding the surface temperature parameter is calculated to be 18.469. Although both models are quite good, the MCSST model is slightly more significant than the NLSST model. It is recognized that the data analyzed in this study are regionally located and represent a small sample relative to global matches typically obtained for AVHRR data. The addition of more matches from much larger areas could possibly generate different results. Such comparisons should be pursued.

b. Daytime

The *GOES-8* matchup dataset was divided into day and night data to regress individual daytime and nighttime equations. This approach takes advantage of the 3.7-μm channel at night and its superior atmospheric transmissivity characteristics relative to the 11- and 12-μm channels (May et al. 1993). These regressions were performed using 9×9 pixel-averaged GOES brightness temperatures. Table 6 shows that regressing a split-window MCSST equation to only the daytime data results in a more accurate SST retrieval than using the 24-h equation. The daytime equation is presented as

$$\text{G8SWD} = 1.0901T_{11} + 1.7339(T_{11} - T_{12}) + 2.8208(\sec\theta - 1) - 299.1844, \quad (6)$$

where G8SWD represents the *GOES-8* split-window daytime SST (°C). The daytime equation demonstrates a bias of −0.01 and an rmsd of 0.79 K against the test data. This outcome is significantly better than the 0.98 K theoretically estimated in Table 3. The Eq. (6) *b* coefficient is significantly smaller in magnitude than the *b* coefficient in Eq. (1), thus the overall sensor noise impact of Eq. (6) is less than previously estimated. The *GOES-8* daytime SST retrieval accuracy is consequently better than anticipated.

c. Nighttime

Split-window, triple-window, and dual-window regressions (using only channels 2 and 4) were performed on the *GOES-8* nighttime matches. The resulting nighttime regression algorithms are presented below:

$$\text{G8SWN} = 1.06T_{11} + 1.8349(T_{11} - T_{12}) + 1.187(\sec\theta - 1) - 289.9502, \quad (7)$$

$$\text{G8TWN} = 1.049T_{11} + 0.8584(T_3 - T_{12}) + 3.1917(\sec\theta - 1) - 286.3231, \quad (8)$$

and

$$\text{G8DWN} = 1.0561T_{11} + 1.1175(T_3 - T_{11}) + 3.5067(\sec\theta - 1) - 287.0664, \quad (9)$$

where G8SWN, G8TWN, and G8DWN represent the *GOES-8* split-window daytime, triple-window nighttime, and dual-window nighttime SST (°C), respectively.

The split-window nighttime equation regressions show an rmsd of 0.84 K for the training set (Table 6). The test set bias of 0.08 K and rmsd of 0.90 K are significantly less accurate than the daytime and 24-h equation results. Application of channel 2 in the MCSST algorithm, however, demonstrates significant improvement over the split-window algorithm results. The triple-window algorithm exhibits an rmsd of 0.75 K for the training set and a bias of 0.01 K and an rmsd of 0.81 K for the test dataset. This outcome is slightly worse than the 0.61 K estimated in Table 3 and cannot be readily explained by equation coefficient differences. Investigation of improved nighttime cloud screening techniques for GOES data could offer potentially better results. The dual-window equation also performs better than the split-window algorithm displaying an rmsd of 0.82 K on the training set and a bias of 0.01 K and an rmsd of 0.87 K against the test dataset. However, the triple-window algorithm is the most accurate nighttime equation.

Regressions of triple-window and dual-window algorithms without the satellite zenith angle term per-

TABLE 7. *GOES-9* SST retrieval regression training and test statistics.

Pixels averaged	Multiple cross correlation	Standard error of estimate (K)	Adjusted R^2
Training set (1667 matches)			
1 × 1	0.9951	0.89	0.9902
3 × 3	0.9959	0.81	0.9919
5 × 5	0.9962	0.78	0.9924
7 × 7	0.9963	0.77	0.9927
9 × 9	0.9964	0.76	0.9929
11 × 11	0.9964	0.76	0.9929
13 × 13	0.9955	0.85	0.9910
15 × 15	0.9949	0.90	0.9899
Pixels averaged		Bias (K)	rmsd (K)
Test set (1660 matches)			
1 × 1		0.01	0.85
3 × 3		0.01	0.79
5 × 5		0.02	0.78
7 × 7		0.01	0.76
9 × 9		0.02	0.75
11 × 11		0.01	0.76
13 × 13		0.01	0.77
15 × 15		0.03	0.82
Equation (10) for various pixel averages (1660 matches)			
1 × 1		−0.01	0.87
3 × 3		−0.01	0.78
5 × 5		−0.01	0.78
7 × 7		0.01	0.76
9 × 9		0.02	0.75
11 × 11		0.01	0.76
13 × 13		0.01	0.77
15 × 15		−0.01	0.82

formed slightly worse than the algorithms that incorporate the term. The RMS differences were 0.05 and 0.09 less accurate, respectively, when the term is ignored. Regression model *t* statistics for coefficients of the satellite zenith angle terms were calculated to be 3.469 and 3.756, respectively, which are statistically significant at the 95% confidence level. These findings again demonstrate better SST retrieval accuracy when the scan angle term is included, albeit an improvement less than 0.1 K.

The *GOES-8* 24-h equation (G8SST) applied to only the daytime test data resulted in a positive bias of 0.09 K and an rmsd of 0.80 K. For nighttime data, the 24-h equation bias was calculated to be −0.02 K with an rmsd of 0.90 K. Comparing these statistics to those in Table 4 concludes that the 24-h equation will be slightly less accurate than utilizing a split-window daytime algorithm in combination with a triple-window nighttime algorithm. Therefore, the latter approach is recommended for the most accurate SST results.

TABLE 8. *GOES-9* SST retrieval algorithm standard error (K).

Algorithm	Training set	Test set
G9SST	0.76	0.75
MCSST without $\sec\theta - 1$	0.76	0.75
MCSST with (T4–T5) $(\sec\theta - 1)$	0.75	0.76
NLSST	0.78	0.76

6. *GOES-9* SST algorithm results

a. 24 hours

As with the *GOES-8* data, regression of the split-window MCSST equation to the entire 24-h dataset for *GOES-9* indicates that a 9 × 9 pixel average provides the most accurate SST results (Table 7). A standard error of 0.76 K is achieved for the training set. A bias of 0.02 K and an rmsd of 0.75 K is achieved for the test set. These findings improve upon those obtained for *GOES-8* and are attributed to the relatively lower NEΔT levels of the *GOES-9* imager channels. The best-fit 24-h regression algorithm is presented below:

$$\text{G9SST} = 1.0083T_{11} + 2.1579(T_{11} - T_{12}) + 0.915(\sec\theta - 1) - 274.9501, \quad (10)$$

where G9SST represents the 24-h *GOES-9* SST (°C).

The results of utilizing Eq. (10) on other pixel array sizes is presented in Table 7. Essentially, scatter increases 0.05 K for a 3 × 3 sample average and 0.08 K for a 15 × 15 sample average. The statistics presented in Table 7 demonstrate that Eq. (10) is the most accurate 24-h *GOES-9* algorithm and that it can also be used for various sample size arrays without significant addition of error.

Table 8 shows that the algorithm neglecting the *GOES-9* satellite zenith angle term performs slightly worse than algorithms that incorporate it. The regression model satellite zenith angle coefficient *t* statistic is calculated to be 3.816, which is statistically significant at the 95% confidence level. This result rejects the hypothesis that the satellite zenith angle coefficient is zero and supports the same results obtained from *GOES-8* data. Regression of a split-window equation following the form of Eq. (1) performs slightly better than Eq. (10) for goodness of fit to the training set. This outcome differs from that obtained from *GOES-8* data and may be due to the relatively lower NEΔT levels of the *GOES-9* IR channels and/or larger number of matchup samples. Equation (10), however, performs best against the test dataset.

The *GOES-9* NLSST algorithm statistics are found not to be as accurate as the MCSST algorithm statistics. This result supports the statistics obtained from *GOES-8* data. Each of the algorithms derived to generate Table 8 were found to be quite good models. However, G9SST is more statistically accurate relative to the test dataset. Additional matches obtained from other regions should be pursued to see if sample location and quantity generate different results.

TABLE 9. *GOES-9* daytime and nighttime algorithm statistics.

Dataset	Bias (K)	rmsd (K)
Daytime split		
Training (945 matches)	0	0.67
Test (950 matches)	0.0	0.65
Nighttime split		
Training (722 matches)	0	0.71
Test (710 matches)	0.02	0.72
Nighttime triple		
Training (722 matches)	0	0.57
Test (710 matches)	−0.02	0.59
Nighttime dual		
Training (722 matches)	0	0.63
Test (710 matches)	−0.01	0.63

b. Daytime

The *GOES-9* matchup dataset was divided into day and night data to regress individual daytime and nighttime equations. Table 9 points out that the *GOES-9* daytime equation is significantly more accurate than the 24-h equation. The daytime split-window algorithm takes the form of

$$\text{G9SWD} = 1.0319T_{11} + 1.9488(T_{11} - T_{12}) + 1.4787(\sec\theta - 1) - 281.8739, \quad (11)$$

where G9SWD represents the *GOES-9* split-window daytime SST (°C). The daytime split-window algorithm shows a bias of 0.01 K and an rmsd of 0.65 K against the test data. This result is slightly better than the 0.73 K estimate from Table 3. The *b* coefficient for Eq. (11) is smaller than the *b* coefficient in Eq. (1) previously used to estimate the error. Thus, channel noise error is not as large as theoretically anticipated.

c. Nighttime

Utilization of the *GOES-9* channel 2 at night demonstrates even further accuracy improvement (Table 9). The triple-window algorithm exhibits a bias of −0.02 K and a remarkable rmsd of 0.59 K against the test data. This outcome is comparable to the 0.52-K estimate from Table 3. The dual-window algorithm also does quite well, demonstrating a bias of −0.01 K and an rmsd of 0.63 K against the test data. The nighttime split-window algorithm bias is 0.02 with an rmsd of 0.72. The *GOES-9* nighttime algorithms are presented below:

$$\text{G9SWN} = 1.0102T_{11} + 2.2097(T_{11} - T_{12}) + 0.9291(\sec\theta - 1) - 275.2568, \quad (12)$$

$$\text{G9TWN} = 0.9845T_{11} + 0.8132(T_3 - T_{12}) + 0.8309(\sec\theta - 1) - 266.6662, \quad (13)$$

and

$$\text{G9DWN} = 0.9853T_{11} + 1.155(T_3 - T_{11}) + 0.8057(\sec\theta - 1) - 266.0039, \quad (14)$$

where G9SWN, G9TWN, and G9DWN represent the *GOES-9* split-window nighttime, triple-window nighttime, and dual-window nighttime SST algorithms (°C), respectively. Table 9 reveals that utilization of channel 2 in the MCSST algorithm provides significant accuracy improvement over the split-window algorithm. Table 9 also demonstrates that using a split-window daytime algorithm in combination with a triple-window nighttime algorithm is significantly more accurate than utilizing a split-window 24-h equation.

7. Summary

GOES-8 and *GOES-9* IR channel brightness temperatures have been matched to moored and drifting buoy SST measurements to derive and evaluate empirical SST algorithms. Satellite and buoy data were collected between November 1995 and April 1996 from digital full disk hemispheric scenes so that a diverse sample of SST magnitudes, atmospheric conditions, and satellite scan angles were encountered. Each image dataset was geolocated, calibrated, and cloud screened with the closest cloud-free pixels matched to buoy SST measurements within 4 h and 25 km.

A linear regression analysis of these matches was performed to generate empirical SST algorithms from the *GOES-8* and *GOES-9* channel brightness temperatures. The *GOES-8* and *GOES-9* IR scanner utilizes a two-detector array that generates some scan striping in the imagery. Investigation of different sample array averages revealed that a 9 × 9 pixel average produces the most accurate SST retrieval results. Averaging less samples increases the data scatter, while averaging more samples also increases the data scatter, potentially due to natural SST changes or gradients within the larger array spatial area. The 9 × 9 sample averaging will impact the user's ability to discern small-scale ocean fronts of potential interest. However, use of 3 × 3 pixel arrays typically generated data scatter less than 0.05 K higher than the 9 × 9 pixel array results. Single-pixel SST processing is not recommended if accurate results are desired. Investigation of a processing filter to remove scan striping should be pursued prior to single-pixel SST processing.

Both *GOES-8* and *GOES-9* indicated that the combination of a daytime split-window algorithm and a triple-window nighttime algorithm statistically outperforms a 24-h split-window algorithm. The best *GOES-8* algorithms exhibit a bias near 0.0 K and rmsd errors of approximately 0.79 K (daytime) to 0.81 K (nighttime) relative to buoy SST measurements. The best *GOES-9* algorithms exhibit a bias near 0.0 K and rmsd errors of about 0.65 (daytime) to 0.59 K (nighttime).

The *GOES-9* results are quite encouraging because they are relatively comparable to *NOAA-14* AVHRR rmsd accuracies of 0.5–0.55 K. The less accurate *GOES-8* performance is thought to be due to its higher IR channel NEΔT values relative to the *GOES-9* imager.

Together, these systems demonstrate that accurate satellite SST retrieval capability is possible from the latest generation of geostationary satellites.

Acknowledgments. The authors appreciate the important efforts of Jeff Hawkins and Joe Turk of NRL-Monterey, who provided the *GOES-8* and *GOES-9* digital datasets that made this study possible; Michelle Parmeter of NAVOCEANO who provided the moored and drifting buoy SST measurements that were matched to the satellite imagery; and Steve Madona of Lockheed Martin for his data processing, algorithm regression, and comparative statistical calculations. This work was sponsored by the Chief of Naval Operations (OP-096) under Program Element 63704N, Satellite Ocean Tactical Applications Program, with the Space and Naval Warfare Systems Command as program manager.

REFERENCES

Anding, D., and R. Kauth, 1970: Estimation of sea surface temperature from space. *Remote Sens. Environ.,* **1,** 217–220.

Barnston, A., and C. Ropelewski, 1992: Prediction of ENSO episodes using canonical correlation analysis. *J. Climate,* **5,** 1316–1345.

Barton, I., 1995: Satellite-derived sea surface temperatures: Current status. *J. Geophys. Res.,* **100,** 8777–8790.

——, A. Zavody, D. O'Brien, D. Cutten, R. Saunders, and D. Llewellyn-Jones, 1989: Theoretical algorithms for satellite-derived sea surface temperatures. *J. Geophys. Res.,* **94,** 3365–3375.

Bates, J., and W. Smith, 1985: Sea surface temperature: Observations from geostationary satellites. *J. Geophys. Res.,* **90,** 11 609–11 618.

Baucom, J., and M. Weinreb, 1996: Characteristics of E/W stripes in infrared images from the *GOES-8* imager. *SPIE Proc.,* **2812,** 587–595.

Clancy, R. M., J. Harding, K. Pollak, and P. May, 1992: Quantification of improvements in an operational global-scale ocean thermal analysis system. *J. Atmos. Oceanic Technol.,* **9,** 55–66.

Dudhia, A., 1989: Noise characteristics of the AVHRR infrared channels. *Int. J. Remote Sens.,* **10,** 637–644.

Emanuel, K., 1988: Toward a general theory of hurricanes. *Amer. Sci.,* **76,** 371–379.

Gallacher, P., R. Rotunno, and K. Emanuel, 1989: Tropical cyclogenesis in a coupled ocean–atmosphere model. Preprints, *18th Conf. on Hurricanes and Tropical Meteorology,* San Diego, CA, Amer. Meteor. Soc., 121–122.

Harris, A., and M. Saunders, 1996: Global validation of the along-track scanning radiometer against drifting buoys. *J. Geophys. Res.,* **101,** 12 127–12 140.

Hogan, T., and T. Rosmond, 1991: The description of the navy operational global atmospheric prediction system's spectral forecast model. *Mon. Wea. Rev.,* **119,** 1786–1815.

Komajda, R. J., and K. McKenzie, 1994: An introduction to the GOES I-M Imager and Sounder instruments and the GVAR retransmission format. NOAA Tech. Rep. NESDIS 82, 56 pp. [Available from the National Oceanic and Atmospheric Administration, U.S. Department of Commerce, Washington, DC 20233.]

Maul, G., 1981: Application of GOES visible-infrared data to quantifying mesoscale ocean surface temperature. *J. Geophys. Res.,* **86,** 8007–8021.

——, and M. Sidran, 1972: Comment on ''Estimation of sea surface temperature from space'' by D. Anding and R. Kauth. *Remote Sens. Environ.,* **3,** 165–169.

——, P. W. deWitt, A. Yanaway, and S. Baig, 1978: Geostationary satellite observations of Gulf Stream meanders: Infrared measurements and time series analysis. *J. Geophys. Res.,* **83,** 6123–6131.

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

——, J. Hawkins, and R. Pickett, 1993: Detecting Gulf of Mexico oceanographic features in summer using AVHRR channel 3. *J. Atmos. Oceanic Technol.,* **10,** 64–75.

McClain, E. P., 1989: Global sea surface temperatures and cloud clearing for aerosol optical depth estimates. *Int. J. Remote Sens.,* **10,** 763–769.

——, W. Pichel, and C. Walton, 1985: Comparative performance of AVHRR-based multichannel sea surface temperatures. *J. Geophys. Res.,* **90,** 11 587–11 601.

McMillin, L., 1975: Estimation of sea surface temperatures from two infrared window measurements with different absorption. *J. Geophys. Res.,* **80,** 5113–5117.

——, and D. Crosby, 1984: Theory and validation of the multiple window sea surface temperature technique. *J. Geophys. Res.,* **89,** 3655–3661.

Menzel, P., and J. Purdom, 1994: Introducing GOES–I: The first of a new generation of geostationary operational environmental satellites. *Bull. Amer. Meteor. Soc.,* **75,** 757–781.

Minnett, P., 1991: Consequences of sea surface temperature variability on the validation and applications of satellite measurements. *J. Geophys. Res.,* **96,** 18 475–18 489.

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

Schwalb, A., 1978: The TIROS–N/NOAA A–G satellite series. NOAA Tech. Memo. NESS 95, 75 pp. [Available from the National Oceanic and Atmospheric Administration, U.S. Department of Commerce, Washington, DC 20233.]

——, 1982: Modified version of the TIROS-N/NOAA A–G satellite series (NOAA E-J). Advanced TIROS-N (ATN). NOAA Tech. Memo. NESS 116, 23 pp. [Available from the National Oceanic and Atmospheric Administration, U.S. Department of Commerce, Washington, DC 20233.]

Smith, E., J. Vazquez, A. Tran, and R. Sumagaysay, cited 1996: Satellite-derived sea surface temperature data available from the NOAA/NASA Pathfinder program. [Available online at http://www.agu.org/eos_elec/95274e.html.]

Strong, A., and E. P. McClain, 1984: Improved ocean surface temperatures from space—Comparisons with drifting buoys. *Bull. Amer. Meteor. Soc.,* **65,** 138–142.

Walton, C., E. P. McClain, and J. Sapper, 1990: Recent changes in satellite-based multi-channel sea surface temperature algorithms. *Paper presented at Science and Technology for a New Oceans Decade,* Washington, DC, Mar. Technol. Soc., 245–250.