

# Overview of the NOAA/NASA advanced very high resolution radiometer Pathfinder algorithm for sea surface temperature and associated matchup database

K. A. Kilpatrick, G. P. Podestá, and R. Evans

Rosenstiel School of Marine and Atmospheric Science, University of Miami, Miami, Florida

**Abstract.** The National Oceanic and Atmospheric Administration (NOAA)/NASA Oceans Pathfinder sea surface temperature (SST) data are derived from measurements made by the advanced very high resolution radiometers (AVHRRs) on board the NOAA 7, 9, 11, and 14 polar orbiting satellites. All versions of the Pathfinder SST algorithm are based on the NOAA/National Environmental Satellite Data and Information Service nonlinear SST operational algorithm (NLSST). Improvements to the NLSST operational algorithm developed by the Pathfinder program include the use of monthly calibration coefficients selected on the basis of channel brightness temperature difference ( $T_4 - T_5$ ). This channel difference is used as a proxy for water vapor regime. The latest version (version 4.2) of the Pathfinder processing includes the use of decision trees to determine objectively pixel cloud contamination and quality level (0–7) of the SST retrieval. The 1985–1998 series of AVHRR global measurements has been reprocessed using the Pathfinder version 4.2 processing protocol and is available at various temporal and spatial resolutions from NASA’s Jet Propulsion Laboratory Distributed Active Archive Center. One of the highlights of the Pathfinder program is that in addition to the daily global area coverage fields, a matchup database of coincident in situ buoy and satellite SST observations also is made available for independent algorithm development and validation.

## 1. Introduction

The need for accurate global sea surface temperature (SST) fields has been receiving increasing attention, primarily because of their importance in understanding variability in the global climate. Satellite SST measurements are attractive because of their global repeated coverage compared to any other type of measurements of this quantity. Space-based multichannel infrared radiometers operating in cloud-free conditions provide the most reliable global SST data sets [Barton, 1995]. Since 1981, the NOAA series of polar-orbiting spacecraft has carried the second generation advanced very high resolution radiometer (AVHRR), an instrument with three infrared (IR) channels suitable for estimating SST [Schwalb, 1978]. These channels are located in the wavelength regions between 3.5 and 3.9  $\mu\text{m}$  and between 10 and 12.5  $\mu\text{m}$ , where the atmosphere is comparatively transparent.

AVHRR-derived SST algorithms have undergone various modifications through the years to improve performance. Details on the evolution of AVHRR SST algorithms are given in section 2. Furthermore, several different radiometers have been flown during the span of AVHRR coverage (at the time of writing, instruments on board NOAA 7, 9, 11, and 14). SST estimates based on data from the various sensors and algorithms were processed on operational time frames, but until recently, there had been no attempt to reprocess the entire database in a consistent manner. To address the need for providing long-term, consistently calibrated global change-related data sets to Earth scientists, the Pathfinder program

was jointly developed by the National Aeronautics and Space Administration (NASA) and the National Oceanic and Atmospheric Administration (NOAA). In addition to providing data sets for global change research, the processing and end-to-end management involved in generating consistent Pathfinder products serve as testing grounds for handling the large volumes of data expected in the next few years from Earth-observing systems.

The objective of this paper is to describe one of the elements of the AVHRR Pathfinder program: the oceans component. The objective of the AVHRR Oceans Pathfinder is to develop a long (>10 years) and consistent time series of global SST fields for climate studies.

The paper is organized as follows. Section 2 describes the evolution of AVHRR SST algorithms that motivated the current formulation of the Pathfinder SST algorithm, introduced in section 3. One of the AVHRR Oceans Pathfinder highlights is that for the first time a large validation database is distributed with the Pathfinder global SST fields to allow interested investigators to develop and validate alternative formulations for the computation of AVHRR-derived SSTs. This database of “matchups” is described in section 4, together with the procedures followed to estimate Pathfinder SST algorithm coefficients. To give potential users of the Pathfinder SST fields a general feel for the performance of the algorithm, we provide an evaluation of the algorithm performance in section 5. Section 6 describes the processing steps involved in the generation of global Pathfinder SST fields. The SST fields may be used for very different purposes, from identifying and tracking specific ocean features to conducting climate-related studies. Such different applications may involve tradeoffs between data coverage and quality (e.g., if finding the Gulf Stream is the goal, one may be more tolerant of lower-quality SST estimates as the

Copyright 2001 by the American Geophysical Union.

Paper number 1999JC000065.  
0148-0227/01/1999JC000065\$09.00

major focus is on SST patterns). For that reason, in section 6 we describe the various quality levels assigned to each pixel in the global SST fields.

## 2. Evolution of SST Algorithms

At IR wavelengths the ocean surface emits radiation almost as a blackbody. In principle, without an absorbing and emitting atmosphere between the sea surface and the satellite it would be possible to estimate SST using a single-channel measurement. In reality, surface-leaving IR radiance is partially attenuated by the atmosphere before it reaches a satellite sensor. Therefore it is necessary to make corrections for atmospheric effects. Water vapor, CO<sub>2</sub>, CH<sub>4</sub>, NO<sub>2</sub>, and aerosols are the major constituents that determine the atmospheric extinction of IR radiance [Minnett, 1990]. Among them, absorption due to water vapor accounts for most of the needed correction [Barton and Checet, 1989].

Several techniques have been proposed to account for the atmospheric absorption of surface IR radiance and to produce accurate retrievals of SST. Anding and Kauth [1970] found that the difference in measurements at two properly selected infrared channels is proportional to the amount of atmospheric correction required. Using differences in brightness temperatures measured by an early satellite radiometer, Prabhakara *et al.* [1974] estimated SST to a reasonable accuracy. In a recent review of techniques to derive SST from satellite IR measurements, Barton [1995] shows that the differential absorption is exploited in all IR SST algorithms and that there is a basic form for most algorithms:

$$\text{SST} = aT_i + \gamma(T_i - T_j) + c, \quad (1)$$

where  $T_i$  and  $T_j$  are brightness temperature measurements in channels  $i$  and  $j$  and  $a$  and  $c$  are constants. The  $\gamma$  term is defined as

$$\gamma = (1 - \tau_i)/(\tau_i - \tau_j), \quad (2)$$

where  $\tau$  is the transmittance through the atmosphere from the surface to the satellite. All AVHRR algorithms share the general form described above, although various modifications have been introduced through the years to improve performance.

In the paragraphs below we briefly present the historical lineage of the different formulations used over the years as the NOAA/National Environmental Satellite Data and Information Service operational algorithm. The multichannel SST (MCSST) algorithm, NOAA's first operational algorithm, was used for several years. The MCSST was based on linear differences in brightness temperatures among AVHRR channels and assumed a constant  $\gamma$  [McClain *et al.*, 1985]. Subsequent improvements by Cornillon *et al.* [1987] incorporated a correction for increased path lengths at larger satellite zenith angles. Other improvements in the atmospheric correction involved nonlinear formulations in which the  $\gamma$  was proportional to the brightness temperatures, as in the cross-product SST algorithm described by Walton [1988] and Walton *et al.* [1990]. The latest version of the operational NOAA algorithm is the nonlinear SST (NLSST) [Walton *et al.*, 1998a] in which  $\gamma$  is assumed to be proportional to a first-guess SST value that can be obtained in various ways.

One important goal of the AVHRR Oceans Pathfinder program was to develop a consensus SST algorithm that mini-

mized the errors or biases in the operational algorithm. The AVHRR Oceans Pathfinder SST algorithm is based on the NLSST algorithm. The NLSST algorithm was chosen as the basis of Pathfinder program by the NOAA/NASA Pathfinder Science Working Group because of the algorithm's operational nature and its familiarity to the scientific user community. The NLSST algorithm published by Walton *et al.* [1998a] has the form

$$\text{SST}_{\text{sat}} = a + bT_4 + c(T_4 - T_5)\text{SST}_{\text{guess}} + d(T_4 - T_5)[\sec(\theta) - 1], \quad (3)$$

where  $\text{SST}_{\text{sat}}$  is the satellite-derived SST estimate,  $T_4$  and  $T_5$  are brightness temperatures in the 10.8 and 11.4  $\mu\text{m}$  bands of AVHRR channel 4 and channel 5, respectively,  $\text{SST}_{\text{guess}}$  is a first-guess SST value, and  $\theta$  is the satellite zenith angle. Coefficients  $a$ ,  $b$ ,  $c$ , and  $d$  are estimated from regression analyses using colocated and coincident in situ and satellite measurements (or matchups). NOAA's operational coefficients are determined and applied separately for day and night. Typically, NOAA produced a set of coefficients using matchups for a certain period; these coefficients would not be modified until there was either a change in the formulation of the operational algorithm or a perceived need (e.g., after the eruption of the Mount Pinatubo volcano in June 1991 or when a new AVHRR was launched). The historical record of NOAA's operational SST algorithms and the associated coefficients can be found in NOAA's Polar Orbiter Data Users Guide Appendix E (<http://www2.ncdc.noaa.gov/docs/podug>). Generally, a single set of day or night operational coefficients is applied for extended periods of time. In contrast, the Pathfinder algorithm uses monthly sets of coefficients for different atmospheric regimes. The same sets of Pathfinder coefficients are applied to both daytime and nighttime data.

SST algorithm coefficients based on IR measurements can be estimated in two major ways. The first alternative involves the use of a radiative transfer model and a set of vertical atmospheric profiles (temperature and humidity), which are used to simulate at-satellite brightness temperatures (BTs). The simulated BTs are subsequently regressed against bottom of the atmosphere in situ temperature measurements from radiosondes (a proxy for SST) in order to derive algorithm coefficients. The use of radiative transfer models requires that the satellite sensor be well characterized and assumes that the set of radiosonde observations used is properly representative of atmospheric conditions. Radiosondes measure only atmospheric temperature and water vapor; however, other factors are also known to impact the radiative transfer process (e.g., air-sea temperature difference, other gases, and aerosols). This semiphysical approach has been adopted, for example, to develop algorithms for the Along Track Scanning Radiometer onboard the ERS-1 satellite [Zavody *et al.*, 1995]. It must be noted that this approach produces so-called "skin temperature" estimates. The skin temperature differs from the "bulk" temperature measured by traditional in situ instruments (e.g., buoys). All IR radiometers, including AVHRR, sense only a skin temperature. However, most oceanographers and climate modelers are interested in a bulk measurement. The skin-bulk temperature difference adds uncertainty to satellite SST retrievals.

A second alternative for estimating SST algorithm coefficients is a regression of the remotely sensed brightness temperatures on in situ buoy temperatures from the matchups.

This statistical approach has been followed to estimate the coefficients for the Pathfinder SST (PFSST) algorithm and produces an SST measurement that has been more closely tuned to a bulk temperature. The differences between skin and bulk SST algorithms, as discussed by *Schluessel et al.* [1990] and *Kearns et al.* [2000], report median values ranging from  $0.1^{\circ}$  to  $0.3^{\circ}\text{C}$  with variations of  $-0.5^{\circ}$  to  $1.8^{\circ}\text{C}$  depending on wind speed and other environmental conditions. The statistical approach to coefficient estimation used in the PFSST algorithm produces a skin temperature biased by the mean skin-bulk temperature difference. This approach has been validated by independent measurements comparing the PFSST with coincident in situ skin measurements with the Marine Atmospheric Emitted Radiance Interferometer (MAERI). *Kearns et al.* [2000] have shown that the average difference between PFSST and MAERI skin temperature measurements was  $0.14^{\circ} \pm 0.31^{\circ}\text{C}$ . This Pathfinder-MAERI difference is comparable to the difference between MAERI and other bulk SST measurements. The use of in situ buoy data in the statistical approach to coefficient estimation allows for errors in both measured BT and measured surface temperature under a variety of conditions and does not require absolute sensor characterization. These types of errors are not included in the model-based approach to coefficient estimation.

One highlight of the AVHRR Pathfinder Oceans program is that for the first time a set of coincident in situ and satellite measurements, which have been used for algorithm development and testing, is being distributed together with the global SST products (see section 4). For the purpose of algorithm coefficient estimation, there are two relevant points regarding the matchups. First, to reduce the variability introduced by space-time separations between the in situ and satellite observations [Minnett, 1991], the Pathfinder matchups have tight space-time constraints: observations are deemed coincident if they occur within  $\pm 30$  min and  $\pm 0.1^{\circ}$  of latitude and longitude of one another. *Schluessel et al.* [1990] showed that the variability in skin-bulk temperature difference were correlated at spatial scales of  $>80$  km, suggesting that matchups should be averaged over larger spatial scales to minimize the skin-bulk temperature difference. The tight space-time constraint in the Pathfinder matchups allows the capture of inherent variations present in both the ocean and the atmosphere, whereas the use of a mean difference on large spatial scales would lose such information. Coefficients derived from the Pathfinder matchups represent the mean state of the oceans and, as such, should account for variations in environmental conditions. Second, the Pathfinder Matchup Database (PFMDB) has been screened carefully to identify most contaminated matchups (e.g., by unidentified clouds), so they could be excluded from the coefficient estimation stage.

### 3. Pathfinder Algorithm

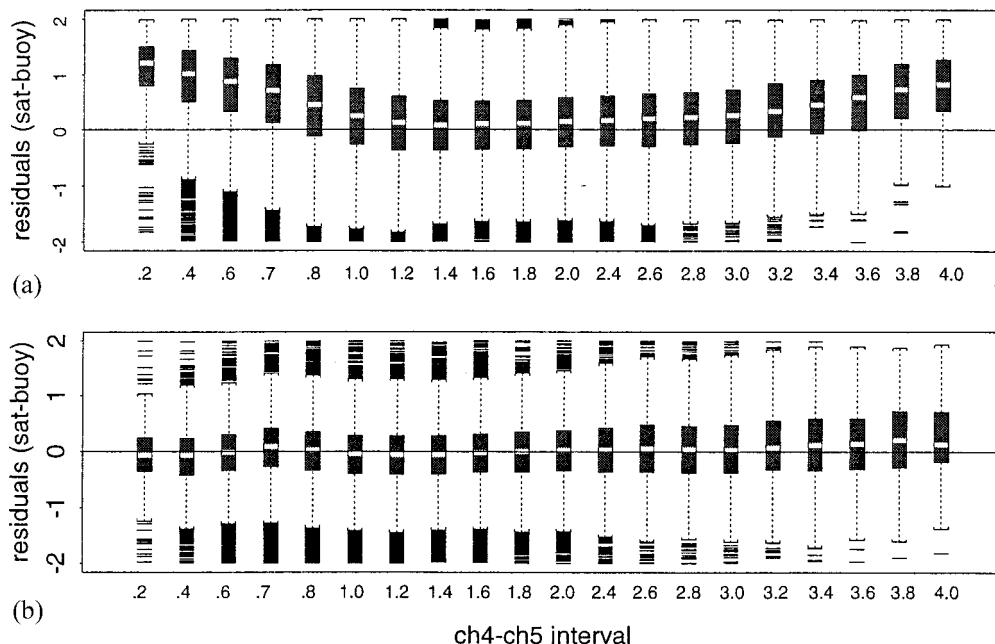
The NLSST formulation developed by C. Walton was adopted as the basis for the AVHRR Oceans Pathfinder SST algorithm because of its adequate performance and its operational nature in NOAA products. Nevertheless, a few minor modifications were introduced to NOAA's NLSST as a result of analysis of the PFMDB; these modifications are described in sections 3.1–3.2.

#### 3.1. Separate Coefficients for Two ( $T_4$ – $T_5$ ) Regimes

Various diagnostics performed on residuals (defined as satellite SST minus in situ SST) from earlier versions of the

PFSST algorithm and matchup database suggested that the association between the ( $T_4 - T_5$ ) values (hereafter referred to as  $T_{45}$ ) and the outcome of the atmospheric correction was significantly different for dry and moist atmospheres. Dry atmospheres are associated with low  $T_{45}$  values, and conversely, moist atmospheres are associated with high  $T_{45}$  values. For instance, there seemed to be a consistent positive bias in SST residuals for  $T_{45} < 0.7^{\circ}\text{C}$ ; this was true for all satellite zenith angle values (Figure 1a). To decrease this bias, we attempted to find optimal empirical transformations to linearize associations between the predicted and predictor variables prior to regression. The estimated empirical transformations showed a change in their shape at  $T_{45}$  values around  $0.7^{\circ}$ – $1.0^{\circ}\text{C}$ . This suggests changes in the underlying functional form of the physical association between  $T_{45}$  and SST. While a full discussion of the possible physical reasons behind this change is quite complex, in the simplest form the balance between various sources of radiance (e.g., ocean and atmosphere) sensed by the AVHRR instrument is likely to have changed as a function of atmospheric moisture. Furthermore, the effects of ocean surface emissivity, air-sea temperature differences, and atmospheric absorbers other than water vapor become more relevant in drier atmospheres. When atmospheric absorption and emission are present, the surface temperature and the brightness temperature no longer coincide. The difference between those two quantities reflects the required amount of atmospheric correction [Barton *et al.*, 1989; Barton, 1992] or temperature deficit (TD). We define the TD as the difference between the in situ SST and the brightness temperature for AVHRR channel 4. As most of the atmospheric effects on infrared radiance can be ascribed to atmospheric water vapor (WV), we examined the association between TD and columnar WV estimates derived from the Special Sensor Microwave Imager (SSM/I) microwave radiometer from records in the PFMDB (Figure 2). As expected, TD increases with WV, although the spread is uniformly large throughout the observed range of WV values. Some of the spread can be tied to the effect of satellite zenith angle, as seen from the trend lines fitted for different satellite zenith angle ranges. The effects of satellite zenith angle were also detected when we examined the association between  $T_{45}$  and WV (Figure 3). Because of the large scatter in the association between water vapor and either TD or  $T_{45}$ , the use of an external WV estimate with its own inherent errors as the sole way of quantifying the necessary amount of atmospheric correction does not seem promising. However, given the general trends observed, we examined how TD varied with  $T_{45}$ , the quantity frequently used as a proxy for atmospheric effects in multispectral approaches. As in the plot of TD versus WV, there seems to be a break in the slope of the fits for satellite zenith angles  $< 45^{\circ}$  at a  $T_{45}$  value of  $0.7^{\circ}\text{C}$  (Figure 4). For satellite zenith angles  $> 45^{\circ}$ , there is no clear break in the fit, although the trend is clearly nonlinear.

As an empirical approach to capturing the change in the nature of the functional association between AVHRR radiances and SST, separate coefficients were estimated for different atmospheric regimes as identified by the  $T_{45}$  value. Algorithm coefficients were estimated separately for (1) low ( $T_{45} < 0.7^{\circ}\text{C}$ ) and (2) intermediate to high  $T_{45}$  values ( $T_{45} \geq 0.7^{\circ}\text{C}$ ). To avoid potential discontinuities in the global PFSST fields at the  $T_{45}$  boundary as the algorithm computation switches from one set of coefficients to another, we implemented a transition in the PFSST calculation. Basically, the procedure in the transitional  $T_{45}$  zone of  $0.5^{\circ}$ – $0.9^{\circ}\text{C}$



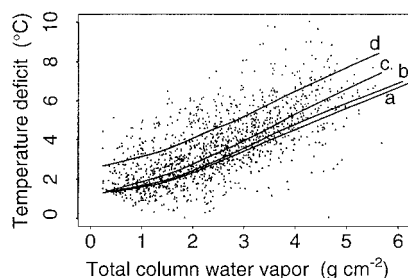
**Figure 1.** Boxplot of the median of the residuals for each  $0.2^{\circ}\text{C}$  interval of channel 4–channel 5 difference from NOAA 11: (a) NLSST NOAA operational algorithm and (b) PFSST algorithm using a piecewise fit as a function of  $T45$ . The white dash in each box represents the median of SST residuals in each  $0.2$  interval of  $T45$ . No results are shown for  $T45$  above 4.2 because of the paucity of matchups above that interval. The box encompasses the central 50% of the residuals (i.e., data between the 25th percentile and the 75th percentile). The whiskers indicate residuals within 1.5 times the width of the box. Individual dashes are extreme outliers (beyond the whiskers' length).

involves the computation of two intermediate PFSSTs using the coefficients corresponding to each  $T45$  regime. The final PFSST is computed as the weighted sum of the two intermediate SSTs, where the weight is a function of the  $T45$  value.

$$\text{PFSST} = w\text{PFSST1} + (1 - w)\text{PFSST2}. \quad (4)$$

Where PFSST is the Pathfinder SST, PFSST1 and PFSST2 are the SSTs computed using the algorithm coefficients for low and high  $T45$  regimes, respectively, and  $w$  is a weighting factor that varies as a function of  $T45$  as follows:

$$\begin{aligned} T45 \leq 0.5^{\circ}\text{C}, \quad w &= 1.0; \\ 0.5^{\circ}\text{C} < T45 < 0.9^{\circ}\text{C}, \quad w &= 1 - [(T45 - 0.5^{\circ}) / (0.9^{\circ} - 0.5^{\circ})]; \\ T45 \geq 0.9^{\circ}\text{C}, \quad w &= 0.0. \end{aligned} \quad (5)$$

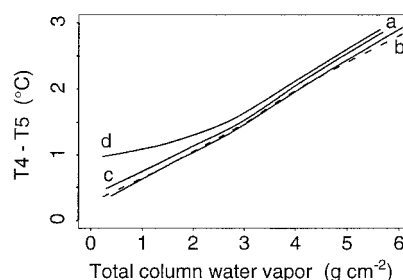


**Figure 2.** TD as a function of integrated SSM/I water vapor values. Nonparametric trend lines are fitted separately for satellite zenith angles of (a)  $0^{\circ}$ – $15^{\circ}$ , (b)  $15^{\circ}$ – $30^{\circ}$ , (c)  $40^{\circ}$ – $45^{\circ}$ , and (d)  $>45^{\circ}$ . One third of the data points, randomly selected, are shown to enhance visualization of trend lines.

That is, for  $T45 \leq 0.5^{\circ}\text{C}$  the PFSST is computed using only the coefficients for low  $T45$  regimes. Similarly, for  $T45 \geq 0.9^{\circ}\text{C}$ , only the coefficients for high  $T45$  regimes are used. For  $T45$  values between  $0.5^{\circ}$  and  $0.9^{\circ}\text{C}$  (i.e., a  $\pm 0.2^{\circ}$  interval around the  $0.7^{\circ}\text{C}$  boundary between  $T45$  regimes) the final SST is a linear combination of the SSTs computed from both sets of coefficients. This approach successfully captures the changes in the functional relationship of SST and  $T45$  (Figure 1b).

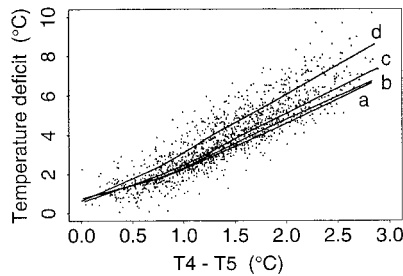
### 3.2. Coefficients Estimated for Monthly Periods Using Temporal Weighting

Earlier versions of the PFSST algorithm used a single set of coefficients per  $T45$  regime for the entire span of an AVHRR's lifetime. Initial diagnostics of the PFMDb, however, suggested the presence of temporal trends in the algo-



**Figure 3.** Difference in brightness temperatures in AVHRR channels 4 and 5 ( $T45$ ) as a function of SSM/I-derived integrated water vapor values. Nonparametric trend lines were fitted separately for satellite zenith angles of (a)  $0^{\circ}$ – $15^{\circ}$ , (b)  $15^{\circ}$ – $30^{\circ}$ , (c)  $40^{\circ}$ – $45^{\circ}$ , and (d)  $>45^{\circ}$ . Points are not shown.





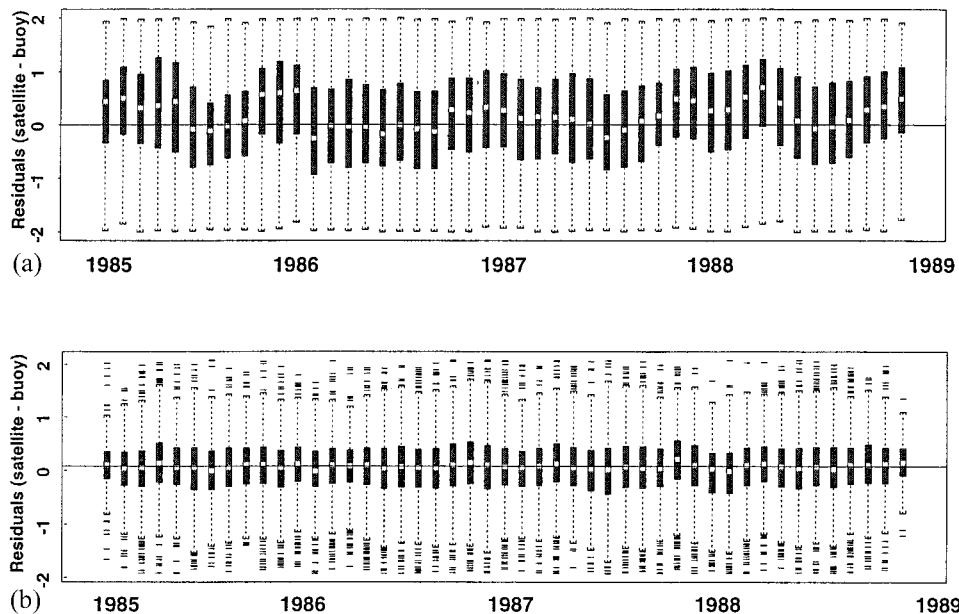
**Figure 4.** Same as for Figure 2, except TD versus difference in brightness temperatures in AVHRR channels 4 and 5.

algorithm performance (Figure 5a). The trends included a variety of temporal scales, from seasonal (e.g., higher bias and rms in SST residuals during Northern Hemisphere summers) to interannual. The interannual trend was of unclear origin and could be due to changes in the radiometric sensitivity of an AVHRR as it aged or to changes in the sensor operating conditions. For instance, during the later stages of NOAA 9's and NOAA-11's operational lifetimes the onboard calibration targets were operated at both a significantly higher temperature than during previous years and a higher temperature than those used for prelaunch sensor characterization, which affected sensor nonlinearity and calibration.

To reduce the presence of trends in the SST estimates, PFSST coefficients starting with algorithm version 3 were estimated on a month-by-month basis. The use of monthly coefficients reduced the bias and eliminated temporal trends in the SST estimates (Figure 5b). Monthly coefficients were estimated using a window of 5 months of matchups centered on the month for which coefficients were being estimated. The use of a 5 month weighted window allows the coefficients to retain

some seasonality and still maximize the space-time distribution of in situ buoy measurements to capture a wide range of atmospheric and environmental conditions. Matchups for each month in the 5 month window were weighted differently: the central month (e.g., month  $N$ ) was assigned a weight of 100%, adjacent months ( $N - 1$  and  $N + 1$ ) were assigned weights of 80%, and the ends of the window (months  $N - 2$  and  $N + 2$ ) were assigned weights of 50%. Because of the lack of data for surrounding months at the start and end of a sensor's lifetime, it is not possible to collect data in a 5 month symmetrical window. For the first and last months of an AVHRR sensor's lifetime we used an asymmetric 3 month window ( $N$ ,  $N + 1$ ,  $N + 2$ ), and for the second and next-to-last months we used an asymmetric 4 month window ( $N - 1$ ,  $N$ ,  $N + 1$ ,  $N + 2$ ). In selecting the temporal weights, no attempt was made to reflect the statistical structure (i.e., temporal correlation) of the SST values. Instead, the main goal was to ensure greater statistical weight for the matchups from the central month. In the latest (version 4.2) Pathfinder coefficients these temporal weights were subsequently combined with robustness weights derived from residuals of a first-estimate of SST values. These robustness weights are in section 4.2 in more detail.

For the Pathfinder algorithm coefficients estimated to date, there were three major exceptions to the general temporal weighting scheme described. First, the NOAA 11 data set (1985–1994) was treated as two separate series, with the separation corresponding to the main eruption of Mount Pinatubo (approximately June 15, 1991). The volcanic eruption of Mount Pinatubo injected large volumes of absorbing aerosols into the atmosphere, dramatically increasing the amount of atmospheric correction required for satellite retrievals. The second exception was the NOAA 9 data used to fill the gap (September 1994 to January 1995) between the demise of



**Figure 5.** Monthly boxplots of residuals for NOAA 9 1985–1989: (a) NLSST NOAA operational algorithm using a single set of coefficients and (b) PFSST algorithm using monthly coefficients for each  $T45$  regime. The white dash in each box represents the median of SST residuals in each month. The box encompasses the central 50% of the residuals (i.e., data between the 25th percentile and the 75th percentile). The whiskers indicate residuals within 1.5 times the width of the box. Individual dashes are extreme outliers (beyond the whiskers' length).

**Table 1.** Percent of 1997 Matchups Provided by Each Agency

Agency	Matchups, %
<i>Moored Buoys</i>	
NDBC	14.6
JMA	0.2
TOGA/TAO	8.6
<i>Drifting Buoys</i>	
AOML	54.0
MEDS	22.6

NOAA 11 and the beginning of NOAA 14's operational period. Because of the short span of this series, a single set of algorithm coefficients was estimated for each T45 regime for the entire NOAA 9 gap, without using temporal weighting. The third exception is the NOAA 7 data; because of the limited geographical and temporal distribution of available in situ buoy data in the PFMDB during the years 1981–1984, coefficients were estimated on an annual basis, without using temporal weighting. Several different temporal weighting schemes were evaluated for use with NOAA 7 by resampling NOAA 14 matchups to emulate the time and space distribution of NOAA 7 (K. A. Kilpatrick, manuscript in preparation, 2001). The results indicate that in the best case, NOAA 7 will have a global median cold bias of  $-0.2^{\circ}\text{C}$ , ranging from  $-0.1^{\circ}$  to  $-0.3^{\circ}\text{C}$  across all latitude bands. This cold bias across all latitude bands is unique to NOAA 7. Further results and discussion on latitudinal biases in the Pathfinder SST for all sensors and years are presented in section 5.

#### 4. Matchup Database and Algorithm Coefficient Estimation

A set of coincident in situ and satellite measurements, which has been used for algorithm development and testing, is being distributed together with the global PFSST products. (A more complete description of the pathfinder matchup database can be found in <http://www.rsmas.miami.edu/groups/rsl/pathfinder/>.) The characteristics of the records contained in any matchup database used for coefficient estimation are extremely important. A matchup database must include observations with a wide geographic and temporal distribution to encompass a broad range of atmospheric and oceanic conditions, as well as observational conditions (e.g., many viewing geometries). Furthermore, the in situ SST measurements should be of high quality. The SST values included in the PFMDB were obtained from buoys deployed and operated by various meteorological and oceanographic agencies or programs in the United States and other countries. The rationale was to use reasonably well calibrated measurements that had undergone some degree of quality control. The in situ SST data originate from two main types of platforms, moored and drifting buoys. The main agencies supplying moored buoy data are the U.S. National Data Buoy Center (NDBC), Japan's Meteorological Agency (JMA), and the Tropical Ocean–Global Atmosphere (TOGA)/Tropical Atmosphere–Ocean (TAO) Project office from NOAA's Pacific Marine Environmental Laboratory. Drifting buoy data were obtained from NOAA's Atlantic and Oceanographic and Meteorological Laboratory (AOML), and the Canadian Marine and Environmental Data Service (MEDS). As an illustration, the 1997

percentage of PFMDB records by source and latitudinal band are provided in Tables 1 and 2. Tables 1 and 2 show that the PFMDB contains a majority of in situ observations from drifting buoys, and both types of buoys are more heavily located in the Northern Hemisphere. Because of the uneven global distribution of buoys, statistical analysis based solely on the PFMDB residuals may not completely capture some latitudinal and seasonal biases.

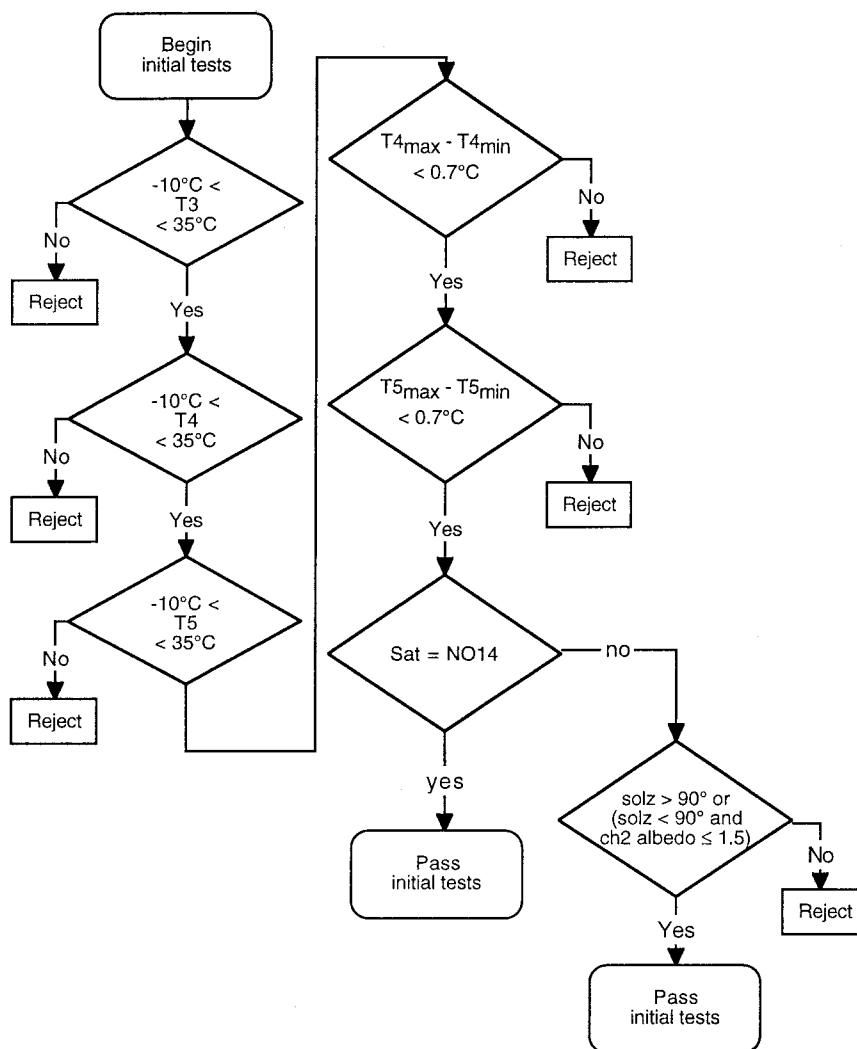
A large number of matchup records in the PFMDB have a wide variety of known problems that tend to result in algorithm failure. Problem conditions include cloud contamination, high atmospheric aerosol or water vapor loading, sensor digitization errors, and slightly miscalibrated in situ buoys. A set of tests is implemented in the PFMDB to identify records likely to display these conditions. We cannot distinguish from the satellite data or the SST data alone which problem condition is present for any given record; therefore the goal of these tests is to eliminate possible suspect matchups. These tests include two stages: (1) a set of initial tests common to all satellites (Figure 6) and (2) a decision tree involving tests derived separately for data from each sensor (Figure 7, only NOAA 14 shown). The nodes and branches of the decision trees developed for each sensor generally follow a common theme of channel differences and threshold tests that partition the data into “suspect condition/cloudy” and “condition free.” Slight variations in ordering of a particular partition test and subtle differences in the absolute values of thresholds for a particular test are present in decision trees for different sensors. This is expected as no two sensors have the same response and calibration. Furthermore, tree partitioning is an empirical technique and is influenced by the data used to “grow” the tree.

Several methods have been proposed in the literature to identify cloud-covered pixels in AVHRR imagery; a few examples include the work of *Saunders and Kriebel* [1988], *Derrien et al.* [1993], *Luo et al.* [1995], and *Cayula and Cornillon* [1996]. Generally, cloud-clearing algorithms use a variety of both visible and IR channels in conjunction with spectral and spatial variability tests to detect clouds. The initial tests used in the Pathfinder project include absolute thresholds for brightness temperature or radiance values in channels that intend to exclude very anomalous values, which may result from digitizer errors. Spatial uniformity tests are intended for further cloud identification. These spatial tests involve thresholds for differences between the minimum and maximum values within a  $3 \times 3$  pixel extraction box for a given channel.

Satellite measurements passing the initial stage 1 tests are subjected to additional screening techniques in stage 2 based on tree models, discussed in paragraphs below in more detail. Earlier versions (version 18) of the PFMDB did not make use of tree models to develop the condition tests but instead relied on filters that had been defined after extensive interactive

**Table 2.** Percent of 1997 Matchups by Latitudinal Band

Latitudinal Band	Matchups, %
$<50^{\circ}\text{S}$	1.6
$50^{\circ}\text{--}25^{\circ}\text{S}$	18.6
$25^{\circ}\text{S--}0$	14.6
$0\text{--}25^{\circ}\text{N}$	18.9
$25^{\circ}\text{--}50^{\circ}\text{N}$	37.0
$>50^{\circ}\text{N}$	9.3

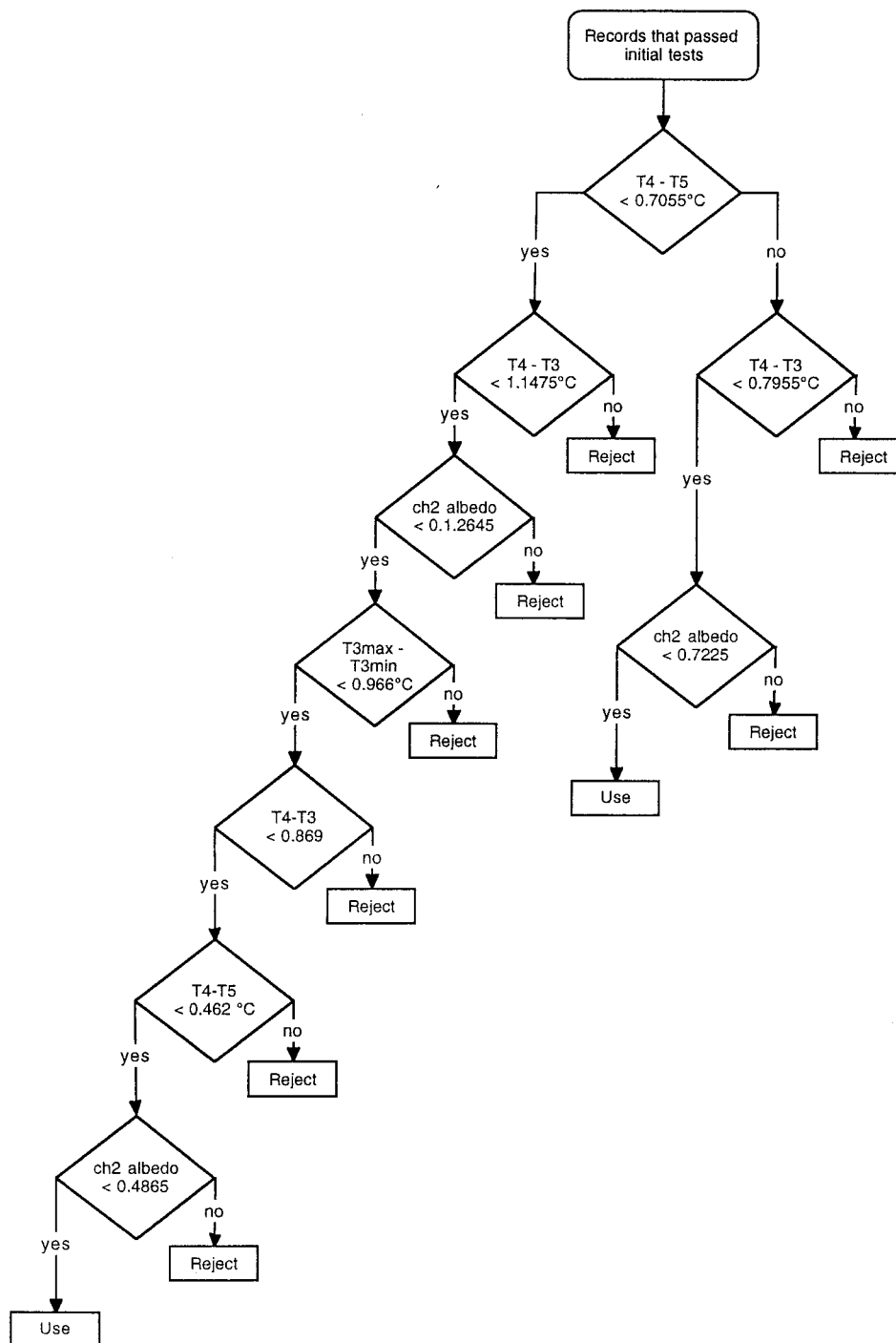


**Figure 6.** Initial common tests used for identification of gross cloud contamination. These tests are applied to all sensors and years.

examination of the data. Although the earlier filters fulfilled their objective of excluding suspect matchups from the algorithm coefficient estimation process, they were overly conservative, rejecting a large amount of potentially usable matchups. For example, in the case of NOAA 9 the earlier filtering method only passed 7.9% of the matchups (8873 out of 102,583 records). Furthermore, selection of tests was done manually after extensive examination of the data; as a result, selection may not have been completely objective. For these reasons a new methodology was developed in version 19 of the PFMDB for stage 2 of the condition/cloud-flagging step prior to coefficient estimation for the version 4.0 algorithm. This new second stage methodology is based on decision tree models described by *Clark and Pregibon [1992]* and *Venables and Ripley [1999]*. There are two main advantages in the new method: tests are selected objectively, and the number of potentially useful matchups is increased. Again, using NOAA 9 as an example, use of a tree-based classification model increased the number of usable matchups by 63%; 13,260 records passed the decision tree model as compared to 8873 with the historical filters.

Tree models are based on binary recursive partitioning

whereby a data set is successively split into increasingly homogeneous subsets. In the present context, tree models can find the best way to predict membership in one of two groups (condition/cloud-contaminated or condition free) as a function of a set of predictor variables that may contain information about the contamination (e.g., differences in brightness temperatures between channels). A tree model is grown from a training sample; in most cases we used a random sample of approximately one third of the matchups passing the initial stage 1 tests, for which the actual classification of all records is known. Matchups in the training set are defined as “bad” or “good,” indicating whether they are potentially contaminated or not. Good matchups in the training set were defined as those with absolute values of SST residuals lower than  $2^{\circ}\text{C}$ . The performance of the resulting tree model developed from the training set is then evaluated using an independent validation set of matchups. Generally, the overall misclassification rate of any of the resulting Pathfinder decision tree models is about 8% of matchups passing the initial test. This overall misclassification rate includes both bad records classified as good and vice versa. Examination of only the records classified as good by the decision trees shows that about 4% have resid-



**Figure 7.** Decision tree for NOAA 14 for years 1996 to present. These tests are developed from a training subset of matchup database records by binary recursive partitioning. Decision trees for other AVHRR sensors are not shown.

uals with absolute values  $> 2^{\circ}\text{C}$ . For example, in NOAA 14 1998, of the 33,397 records in the PFMDB classified as good by the decision tree, 1577 (4.7%) of these records had large residuals. For comparison, about 6% of the records passing the historical filtering had large residuals. These statistics indicate that the use of a decision tree model increased the number of available matchups and their quality. The failure of the decision trees to identify these few remaining matchups with high residuals may indicate problems with the in situ values (e.g., a miscalibrated buoy), anomalous atmospheric conditions,

small-scale cloud, or other inherent limitations in the current algorithm. These outliers (matchups with high SST residuals) can unduly influence coefficient estimates and need to be excluded from the estimation procedure. In coefficient estimation for versions prior to version 4.0 we had excluded matchups with an absolute value of residuals  $> 2^{\circ}\text{C}$  with respect to a first-guess SST typically based on the operational NLSST or the prior years' Pathfinder coefficients. We explored other alternatives that would not involve a fixed threshold for exclusion of matchups showing high SST residuals to limit the im-



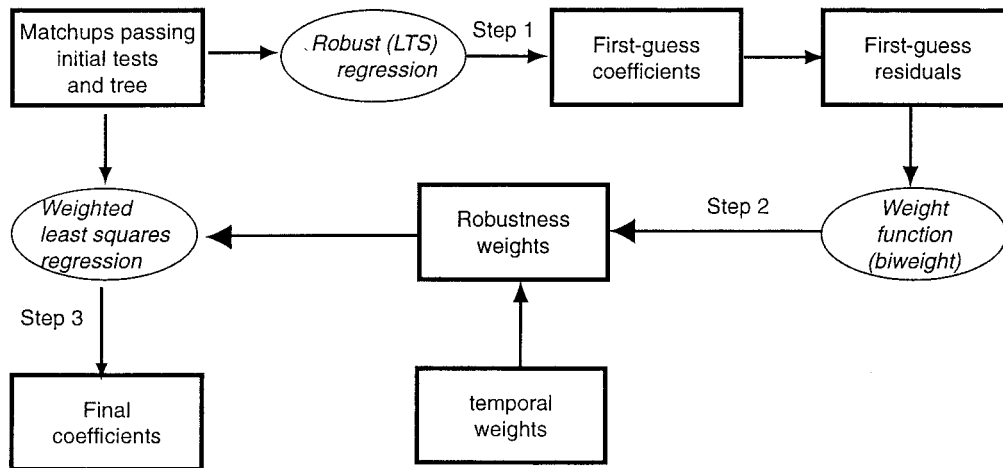


Figure 8. Flow diagram of steps in coefficient estimation procedure using modern resistant regressions.

part of outliers and miscalibrated buoys. Such a procedure was implemented in version 4.0 of the Pathfinder products by use of modern resistant regression. Resistant regression procedures such as the least trimmed squares (LTS) [Rousseeuw and Leroy, 1987] return a regression estimate that minimizes the sum of the smallest half of the squared residuals. LTS methods have a very high breakdown point. In statistical terms the breakdown point indicates the proportion of outliers that can be present in a data set before estimates are strongly influenced; the higher the breakdown point, the more resistant the procedure [Lanzante, 1996].

The estimation of SST algorithm coefficients involves three steps. First, resistant regression is used to obtain first-guess coefficients and first-guess SST residuals. Second, robustness weights are computed using the first-guess SST residuals to exclude large outliers. Finally, a weighted least squares regression determines the final coefficients. Figure 8 presents a flow diagram of the coefficient estimation procedures.

#### 4.1. Estimation of First-Guess Coefficients and SST Residuals

All matchups that passed the suspect condition/cloud tests were used to estimate a preliminary (first-guess) set of algorithm coefficients using an LTS regression. The LTS resistant regression was repeated for each *T45* regime and for each period for which algorithm coefficients were estimated. The preliminary algorithm coefficients obtained from the LTS were used to compute a first-guess SST and first-guess SST residuals.

#### 4.2. Computation of Robustness Weights

The goal of this step is to assign reduced weights to records with large first-guess residuals (for instance, those due to unidentified cloud contamination) in order to reduce their influence on final coefficient estimation. To derive robustness weights from the first-guess SST residuals, we first estimated for each period and *T45* regime the median of the absolute values of first-guess residuals; this quantity is the median absolute of deviations (MAD). Second, we used the bisquare function  $B(u)$  to compute robustness weights. The bisquare function is commonly used in robust regressions [Cleveland, 1993], but other functions could have been chosen (e.g., tricube). The only requirements in choosing a function are that

it be (1) nonnegative and (2) symmetrically decreasing from its central value. The bisquare function of argument  $u$  has the following form:

$$B(u) = (1 - u^2)^2, \quad |u| < 1; \\ B(u) = 0, \quad |u| \geq 1 \quad (6)$$

The first-guess residuals (denoted as  $\varepsilon$ ), their corresponding MAD (for a given period and *T45* regime), and the bisquare function were used to compute robustness weights  $r$  as

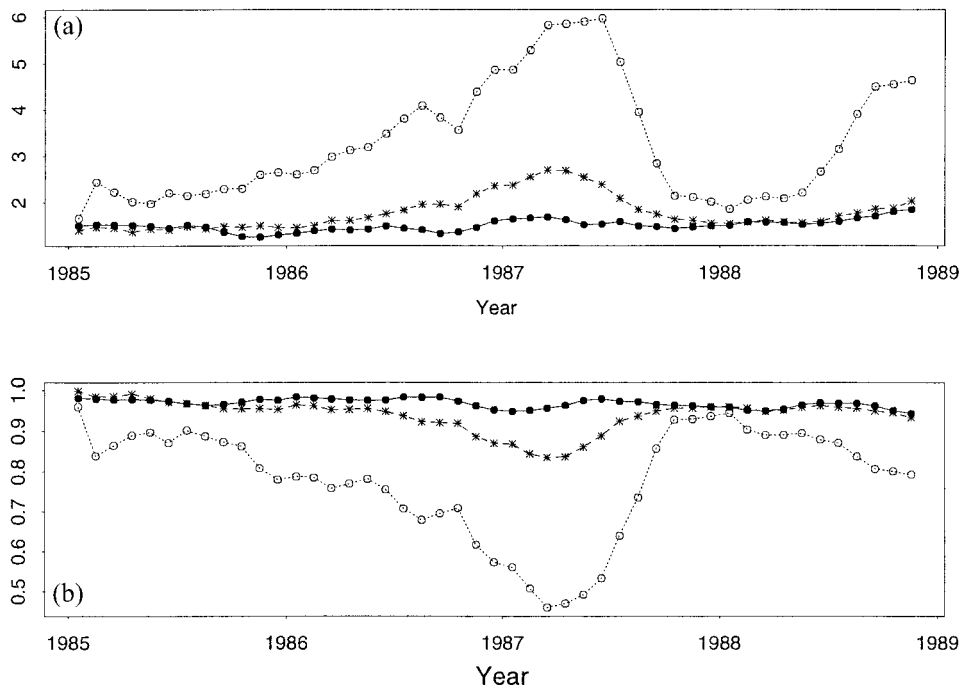
$$r = B[\varepsilon/(6\text{MAD})]. \quad (7)$$

Equation (7) indicates that the robustness weights have a value of zero for matchups with  $|\varepsilon| > 6\text{MAD}$ . The factor of 6 multiplying the MAD was selected so that if the first-guess residuals have an underlying Gaussian distribution, this threshold would be approximately equivalent to rejecting first-guess residuals beyond  $\pm 4$  standard deviations. In most cases, MAD values ranged between  $0.3^\circ$  and  $0.4^\circ\text{C}$ ; this implies that residuals with absolute values  $> \sim 1.8^\circ$  were excluded (i.e., had weights equal to zero).

#### 4.3. Estimation of Final Coefficients

The last stage of the coefficient estimation procedure involves a weighted least squares procedure. The robustness weights derived in the previous stage were multiplied by the temporal weights in the 5 month window (described in section 3.2). The resulting values were the final weights used as input to a weighted least squares regression. The weight assigned to a particular matchup during coefficient estimation therefore was a function of both its first-guess SST residual and its temporal separation from the month for which coefficients were being estimated. The coefficients estimated by the weighted least squares regression were used to produce the global Pathfinder version 4.2 SST fields.

What is the advantage of using the approach described in previous paragraphs? Besides the fact that the procedure to identify large outliers is more objective, earlier versions of the PFSST coefficients showed temporal fluctuations at scales of months to years. Although fluctuations in the coefficient for one term of the algorithm usually were compensated somewhat by the values of coefficients for other terms, it was un-



**Figure 9.** Time series of the first two algorithm coefficients for NOAA 9 1985–1988. (a) Coefficient 1 is the constant term, and (b) coefficient 2 is the  $T_4$  multiplier. The coefficients correspond to matchups with  $T_{45}$  values  $> 0.7^\circ\text{C}$ . The open circles indicate coefficient values estimated using unweighted least squares regression on all matchups that passed the cloud-flagging tests (procedure A in text). The asterisks denote coefficients estimated by (1) using straight least squares as described in the previous sentence, (2) excluding residuals with absolute values  $> 2.2^\circ\text{C}$ , and (3) reestimating coefficients using straight least squares (procedure B in text). The solid circles indicate coefficient values estimated via a resistant regression, followed by a weighted least squares regression (procedure C in text).

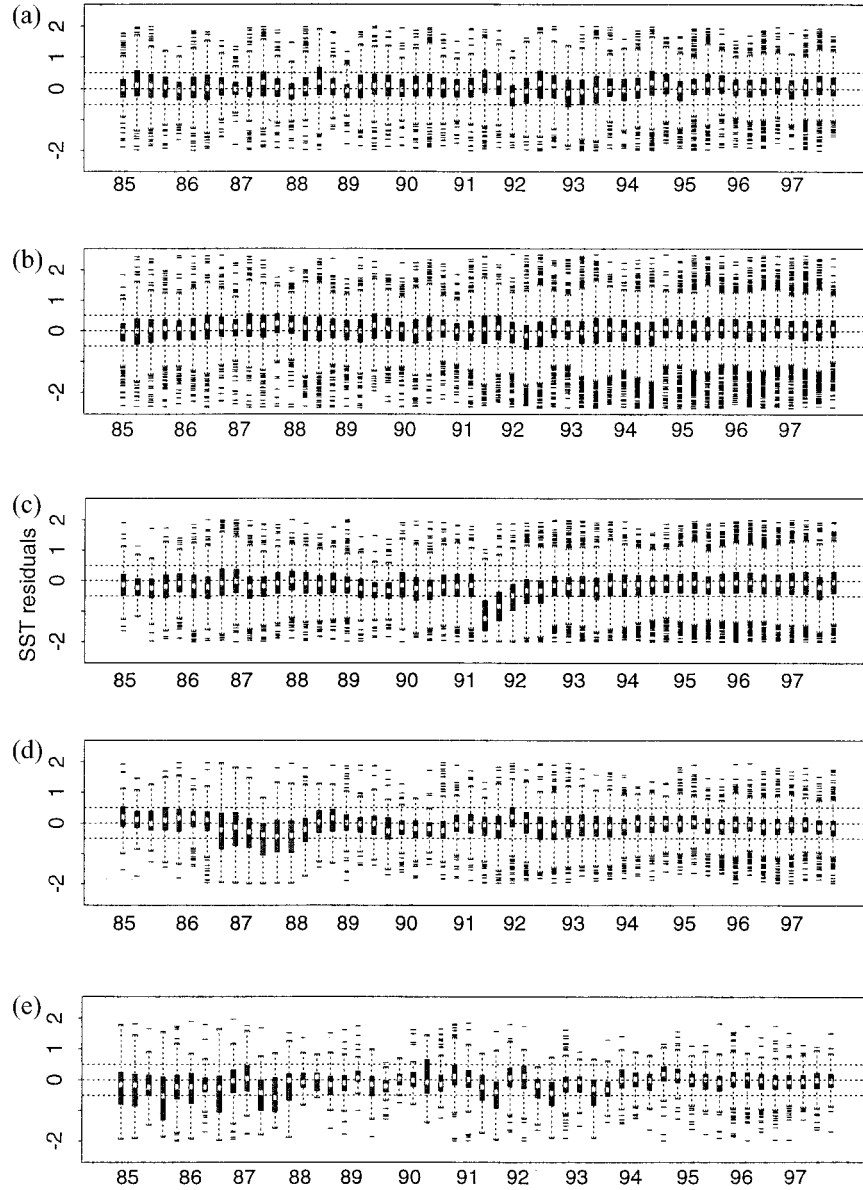
clear if this had consequences on algorithm performance. To assess the advantages of the resistant procedure, we compared coefficients derived by three different procedures. We first estimated a set of coefficients (procedure A) using an ordinary least squares regression on the matchups that passed the cloud-flagging tests. A second set of coefficients (procedure B) was estimated by excluding procedure A residuals  $> \pm 2^\circ\text{C}$  and reestimating coefficients (using ordinary least squares). A third set of coefficients (procedure C) was derived using the new protocol described in sections 4.1–4.3.

Figure 9 shows time series of the first two PFSST algorithm coefficients (the constant term and the term multiplying  $T_4$ ) for NOAA 9 matchups. The three lines correspond to coefficients estimated following procedures A, B, and C, as described above. It is clear that the temporal stability of the procedure C coefficients is much greater. This stability in the coefficients may allow the distribution of a near “real-time” interim Pathfinder SST product before archived in situ data become available and final coefficients are estimated. These results may make one consider whether separate monthly coefficients are necessary. Although the resistant regression estimation seems to have reduced considerably the seasonal fluctuations in the coefficients, they have not disappeared entirely. Furthermore, there are still unexplained low-frequency trends in the coefficient values. Therefore, for the current version of the Pathfinder fields we still use monthly coefficients.

## 5. Algorithm Performance Evaluated From the Matchups

To give potential users of PFSST fields a general feel for the variability in SST estimates, we provide boxplots of SST residuals (PFSST minus in situ SST) for the AVHRRs that operated during 1985–1995 (Figure 10). Boxplots are shown for four latitudinal bands:  $40^\circ\text{--}20^\circ\text{S}$ ,  $20^\circ\text{S--}20^\circ\text{N}$ ,  $20^\circ\text{--}40^\circ\text{N}$ , and  $40^\circ\text{--}60^\circ\text{N}$ . No results are shown for latitudes below  $40^\circ\text{S}$  or above  $60^\circ\text{N}$  because of the paucity of matchups in those regions.

In general, each box and whiskers contains at least 100 matchups (high latitudes, 1985) and as many as 12,000 (mid-latitudes, 1997); therefore the statistics are considered stable. The Pathfinder approach for the most part yields high-quality SST retrievals: for most periods the box that encloses the central 50% of the residuals falls well within the  $\pm 0.5^\circ\text{C}$ . Exceptions include unresolved clouds and aerosols. With respect to all the buoys in the matchup database the PFSST (bias, mean, and median) is within  $0.02^\circ\text{C}$ , and the standard deviation is  $0.53^\circ\text{C}$ . These aggregate global statistics represent all condition/cloud-free records present in the PFMDB from 1985 to 1998 and include all sensors. Aggregate statistics, however, can be misleading; positive residuals in some latitudinal bands are offset by negative residual in others, resulting in a near-zero bias when combined. In addition, the in situ buoys used in the comparison are themselves not calibrated to an accuracy of



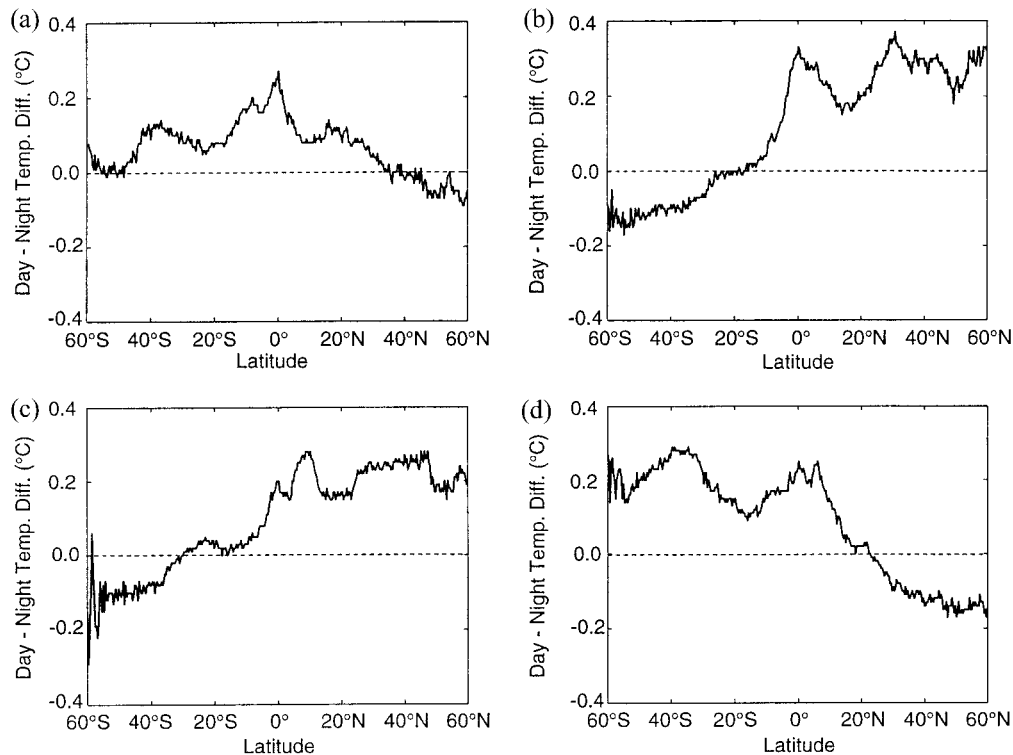
**Figure 10.** Boxplots of PFSST residuals (PFSST minus buoy SST) including NOAA 9 (January 1985 to November 1988), NOAA 11 (November 1988 to September 1994), NOAA 9 gap (September 1994 to March 1995), and NOAA 14 (February 1995 to December 1997) for (a) 40°–60°N, (b) 20°–40°N, (c) 20°–40°S, (d) 40°–60°S, and (e) 40°–60°S. Each box and whiskers in a panel corresponds to a 3 month period. The dashed horizontal lines in each panel indicate  $-0.5^{\circ}$ ,  $0^{\circ}$ , and  $0.5^{\circ}$ .

$0.02^{\circ}\text{C}$ . Nevertheless, in situ buoys are the best currently available standard on a global scale, and results indicate that the PFSST algorithm performs well.

While we cannot determine the absolute accuracy of the PFSST fields, examination of the data by latitudinal band suggests that the median value of the residual is probably closer to  $0.1^{\circ}\text{C} \pm 0.5$ . The rms of residuals in the PFMDB are twice that reported for MAERI/Pathfinder matchups [Kearns *et al.*, 2000], suggesting that the uncertainty in the PFSST is likely to be a combination of both differences in skin-bulk temperatures and buoy calibration. One important feature is the bias introduced by absorbing aerosols from the Mount Pinatubo explosion (June 1991). This is particularly noticeable in the tropical latitudes, and it takes several months for residuals to approach normal. Another important point is that in general, the tropical

band shows a negative bias of about  $0.1^{\circ}$ – $0.2^{\circ}\text{C}$ . That is, the Pathfinder SST algorithm does not produce sufficient correction for atmospheric effects that attenuate the surface signal. In contrast, the band between  $20^{\circ}$  and  $40^{\circ}\text{N}$ , where most matchups occur, tends to show a positive bias: the Pathfinder algorithm overcorrects for atmospheric effects. It is remarkable that the central half of the residuals (denoted by the boxes) shows a fairly tight distribution.

We stress that the matchup residuals used to produce Figure 10 were computed using the actual in situ buoy SST as the first-guess ( $\text{SST}_{\text{guess}}$ ) value in the algorithm. The in situ buoy temperature was chosen as the first guess to provide the greatest accuracy estimate of the algorithm and avoid complicating factors associated with the choice of climatologies used as the first-guess SST in the processing of global fields. In contrast,



**Figure 11.** Seasonal daytime minus nighttime temperature differences versus latitude: (a) January–March, (b) April–June, (c) July–September, and (d) October–December. Data are from NOAA 14 global 36 km Pathfinder SST global area coverage (GAC) weekly files 1995–1997.

the first-guess SST in the Pathfinder field calculations is the Reynolds optimally interpolated (OI) SST and is discussed in section 6. The SST residuals computed using the values in the global fields then can be slightly different. This difference is, however, quite negligible. For NOAA 14 matchups 1995–1997 the difference in the retrieved SSTs using the Reynolds reference as the first guess as opposed to the in situ buoy SST as the first guess is  $0.001^{\circ}\text{C} \pm 0.120$ . Walton *et al.* [1998a] previously demonstrated that the NLSST algorithm is relatively insensitive to errors in the climatological temperature used as the first-guess SST ( $\text{SST}_{\text{guess}}$ ), and by extension the PFSST should behave in the same manner.

As stated earlier, statistical analysis of the PFMDB may not capture accurately some regional and seasonal biases because of the uneven geographical distribution of the buoys. Analysis of weekly global Pathfinder fields as function of season and latitude indicates the presence of a seasonal day-night temperature difference in the retrievals. Figure 11 shows the day-night temperature difference for weekly 36 km Pathfinder fields as a function of latitude during four seasons. At the equator, daytime SST retrievals are consistently warmer than nighttime retrievals. This is due to the fact that the cloud-free conditions suited for AVHRR measurements of SST are also frequently conducive to the generation of diurnal heating effects in the near-surface layer of the ocean [Schuessel *et al.*, 1990]. In the equatorial regions, high solar insulation and low winds can lead to large diurnal heating, with afternoon surface temperatures higher than during the night. The AVHRR daytime overpass is at  $\sim 1430$  LT. A similar geophysical signature is seen at high latitudes in both hemispheres during the summer, again, when solar insulation is high and wind speeds are low. As stated previously, the AVHRR is measuring radiances that have their

origin at the temperature of the skin, but the Pathfinder algorithm produces SSTs biased to the mean global skin-bulk temperature difference. Whenever the skin-bulk temperature difference deviates from the mean global value, day-night differences in the retrievals are likely to occur. Wind speed also plays a role in the magnitude of the skin-bulk temperature difference. In the trade wind latitudes, where higher wind speeds are expected to suppress the growth of diurnal temperature signals, a commensurate decrease is seen in the difference between the day-night retrievals.

An unexpected result seen in Figure 11 is the large negative day-night difference during the winter at high latitudes, indicating warmer nighttime retrievals. It is difficult to envision a geophysical explanation for this result. We suggest that this may be related to onboard calibration issues, specifically in the temperature of the blackbody. At a given latitude the period since the satellite enters eclipse and begins to cool during the nighttime part of each revolution is dependent on the solar declination (i.e., time of year). Given that there are significant temperature changes aboard the spacecraft around the orbit, the dynamic thermal environment in the AVHRR blackbody could give rise to small calibration errors that are a function of the position of the spacecraft relative to entering eclipse [Brown *et al.*, 1985, 1993]. The prelaunch calibrations of the blackbodies are accurate to only  $\pm 0.1^{\circ}\text{C}$ , which is comparable to the day-night differences. Clearly, PFSST accuracy is likely to be dependent on the geographical location and season.

## 6. Processing of Global Pathfinder SST Fields

So far, our discussion has mostly focused on the Pathfinder matchup database and how it is used to estimate SST algorithm



coefficients and determine algorithm performance. We now turn our attention to the processing details involved in the creation of the daily 4 km global area coverage (GAC) PFSST fields. Processing of the version 4.2 Pathfinder AVHRR data involves four steps: ingestion (which includes both conversion to calibrated radiance and georeferencing), Pathfinder SST calculation, spatial binning, and temporal binning. The software used to process the Pathfinder data is the processing system developed at the University of Miami. AVHRR data are received in files containing data from approximately one orbit. Because of its large size, the orbit file is split into six to eight pieces prior to ingestion and scanned to obtain navigation and calibration information needed for processing.

### 6.1. Calibration and Count-to-Radiance Conversion

During ingestion the raw AVHRR digital counts are converted to calibrated radiance values. AVHRR calibration methods have changed through the years, and various groups historically have used different calibration procedures. However, one of the results of the AVHRR Pathfinder project is a consensus AVHRR Pathfinder calibration protocol. Briefly, the goal of NOAA/NASA AVHRR Pathfinder calibration procedure is to minimize spurious trends in the long-term records of AVHRR-derived geophysical products and to ensure homogeneity across geophysical products derived from AVHRR sensors aboard the NOAA 7, 9, 11, and 14 spacecrafts.

In principle, the radiance sensed by an AVHRR IR channel is a function of the temperature of the target (Earth) and the spectral response of the channel. Laboratory calibration of channels 4 and 5 indicated that the detector's response is nonlinear [Brown *et al.*, 1985; Weinreb *et al.*, 1990]. This nonlinearity depends both on the AVHRR operating temperature and on the scene radiance. The procedure to convert AVHRR counts to radiance for the IR channels is a three-step process. The first step is a linear transformation of counts to radiance, based on the onboard blackbody calibration devices (space view and sensor baseplate). The second step applies a nonlinearity correction factor (derived from prelaunch calibration data) to correct radiance values for channels 4 and 5 only. The third and final step involves the use of lookup tables to convert radiance to temperature as a function of the sensor's operating temperature. Readers interested in the details of sensor calibration are referred to Walton [1988], Walton *et al.* [1998b], and Brown *et al.* [1993] for IR channels 3, 4, and 5 and Rao and Chen [1995, 1996] for visible and near-IR channels 1 and 2. Additional information can also be found in section 7.1 of the NOAA KLM User's Guide (<http://www2.ncdc.noaa.gov/docs/klm/index.htm>).

### 6.2. Navigation: Clock and Attitude Correction

Following the conversion of counts to calibrated radiance, the georeferencing information associated with each piece is determined. The position on Earth of any pixel in the AVHRR data can be determined from an orbital model and a sensor model. The orbital model determines the precise location of the satellite along the orbital track, using the time of each processing piece. The sensor model determines the direction in which the AVHRR instrument is pointing. The major sources of error in geolocating AVHRR data are (1) drift in the spacecraft clock, which produces errors in the estimated along-track position, and (2) uncertainty errors in spacecraft and sensor attitude. To minimize these errors, the time was corrected for satellite clock drift, and a nominal attitude correction was

applied. The clock drift rate was determined from a database of satellite clock time and Earth time offsets collected from high-resolution picture transmission passes obtained from NOAA DOMSAT receiving stations at Wallops Island, Virginia, and Fairbanks, Alaska. The nominal attitude was determined by averaging the absolute attitude of the spacecraft over many geographic locations and times along the orbital track. (Readers interested in further details on the methods for determining these corrections, including a listing of the factors for each satellite are referred to [http://www.rsmas.miami.edu/groups/rsl/pathfinder/processing/proc\\_app\\_a.html](http://www.rsmas.miami.edu/groups/rsl/pathfinder/processing/proc_app_a.html)). The resultant navigation information provides the mapping parameters needed to convert the satellite coordinates of pixel and scanline to latitude and longitude.

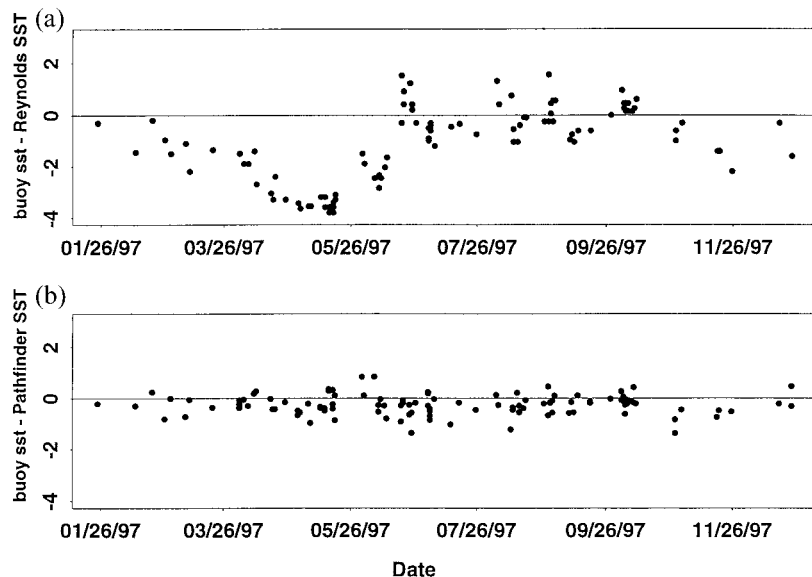
### 6.3. SST Computation

Once the calibrated radiance and navigation information is calculated, the Pathfinder algorithm is applied to compute the SST. The Pathfinder algorithm requires a first-guess SST. The first-guess used in version 4.2 of the Pathfinder GAC product is derived from the optimally interpolated SST weekly National Centers for Environmental Prediction (NCEP) SST analysis of global buoy, ship, and NOAA AVHRR SST data developed by Reynolds and Smith [1994]. To create the first-guess Reynolds reference fields, three consecutive weekly NCEP SST analysis were averaged using a 0.25:0.5:0.25 weighting to create a reference field centered on the middle week as recommended by R. Reynolds (personal communication, 1994; see [http://www.cdc.noaa.gov/cdc/reynolds\\_sst.info.html](http://www.cdc.noaa.gov/cdc/reynolds_sst.info.html)).

### 6.4. Pixel-by-Pixel Science Quality Flags

One of the main goals of the AVHRR Oceans Pathfinder project is to produce global SST fields of a quality as good as possible. Nevertheless, processing errors (cloud flagging and SST algorithm) may result in poor SST estimates. The next step in the processing is to perform a series of tests to assess the likelihood that a pixel contains an SST value of suspect quality. The various tests are then combined to define eight overall quality levels for a pixel. The overall pixel quality levels are taken into account during the spatial binning stage. The outcome is an overall bin quality level. The goal of the pixel-by-pixel quality flags and assignment of an overall quality level is to minimize the occurrence of suspect pixels and still maintain flexibility in the population of acceptable pixels. End users can then make their own decisions on the trade-off between geographic coverage and accuracy on the basis of quality level. The various steps involved are described in subsequent paragraphs. First, a group of nonsequential quality tests is applied on a pixel-by-pixel basis to flag certain conditions. Results of each of the nine individual condition tests are stored in two eight-bit masks and used later to establish the overall quality level of the pixel. Both the two masks and the overall quality level are stored in the final PFSST product file. Users may wish to recombine the mask bits to create alternative data-screening techniques to meet their specific requirements, rather than use the overall quality levels as defined by the project. Each of the nine conditions in the masks is tested independently, and no sequential order is implied by the presentation of the tests.

**6.4.1. BT.** Brightness temperatures for AVHRR channels 3, 4, and 5 must be  $\geq -10^{\circ}\text{C}$  and  $\leq 35^{\circ}\text{C}$ . This test identifies sensor digitizer errors or very cold pixels associated with high cloud tops. BTs outside these thresholds are not geophysically reasonable.



**Figure 12.** Time series (1997) of NDBC buoy 41001 minus (a) Reynolds and (b) PFSST residuals.

**6.4.2. Cloud test.** Pixel must pass a suite of cloud-flagging tests, arranged as a decision tree and defined for each satellite and year (Figures 6 and 7). Details regarding cloud detection tests were provided in section 4.

**6.4.3. Uniformity test 1.** The BTs for channels 4 and 5 are examined for all pixels within a  $3 \times 3$  pixel box centered on the pixel being classified. The difference between maximum and minimum BTs for each channel within this box must be  $<0.7^{\circ}\text{C}$ . This test seeks to identify contamination by small clouds and is based on the assumption that BTs should be relatively uniform at small scales under cloud-free conditions. Analysis of the PFMDB has shown that for channel uniformity thresholds below  $0.7^{\circ}$ , no significant bias was detected in PFSST estimates and the rms of PFSST residuals was relatively uniform. However, pixels in areas of sharp frontal features on scales smaller than 12 km may be identified erroneously as suspect by this test.

**6.4.4. Uniformity test 2.** This test is the same as that described above; however, the minimum/maximum threshold is relaxed to  $1.2^{\circ}\text{C}$ . This higher uniformity threshold allows more pixels to pass the test at the expense of accepting pixels with a slightly colder SST bias.

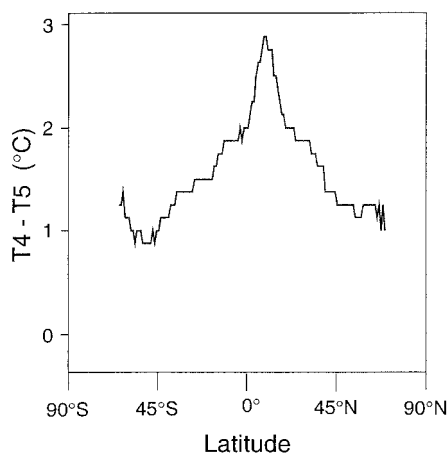
**6.4.5. Zenith angle test 1.** Satellite zenith angle must be  $<45^{\circ}$  to pass this test. At higher zenith angles, radiation emitted by the ocean has to go through a longer atmospheric path before reaching the AVHRR instrument, with consequently higher chances of being attenuated. The received radiance therefore is likely to have a lower proportion of radiance originating from the ocean's surface and a greater proportion of radiance reemitted by the atmosphere. The disadvantage of limiting zenith angles is the loss in geographic coverage. The satellite zenith angle thresholds were determined by examining plots of residuals versus  $T45$  for ranges of satellite zenith angle (not shown) from records in the matchup database. From these plots we identified limits that minimize the likelihood of contaminated pixels while maintaining a reasonable population size of good pixels.

**6.4.6. Zenith angle test 2.** Satellite zenith angle must be  $<55^{\circ}$ . This is similar to zenith angle test 1, but it allows a larger

range of acceptable zenith angle values, with the goal of gaining geographic coverage.

**6.4.7. Reference test.** The absolute difference between the PFSST for the pixel considered and the 3 week Reynolds reference SST field must be  $\leq 2^{\circ}\text{C}$ . Pixels in the Reynolds reference field represent a  $1^{\circ}$  cylindrical equidistant projection of SST in degrees Celsius. Because of the low time and space resolution of the Reynolds reference and the limited geographical coverage of the in situ data used in the Reynolds OI analysis, this reference test may occasionally bias quality flags in coastal areas, regions not monitored with in situ buoy observations (e.g., high latitudes, equatorial Atlantic, and Indonesia) and regions having large SST gradients in time or space. An example of the potential quality bias associated with the use of the Reynolds reference in high-gradient areas is shown in Figure 12. In this example we show NDBC buoy 41001, a 6 m Navy Oceanographic and Meteorological Automatic Device buoy, which is located at  $34.68^{\circ}\text{N}$ ,  $72.64^{\circ}\text{W}$ , 280 km east of Cape Hatteras near the Gulf Stream. Time series of buoy SST minus Reynolds OI SST (Figure 12a) and buoy SST minus PFSST (Figure 12b) show that during April and May of 1997 the Reynolds reference was more than  $2^{\circ}\text{C}$  warmer than the buoy observations, while the PFSST retrieval shows almost no bias. In this April/May time period, Pathfinder pixels around this buoy are flagged as suspect (quality level 3 or less; see later Figure 14) because of extreme departure from the Reynolds reference.

**6.4.8. Stray sunlight test.** An examination of data stratified by satellite zenith angle and side of the AVHRR scan line (left and right of nadir) revealed potential problems under certain Earth-Sun-satellite configurations. We have found that stray sunlight contamination of pixels occurs primarily on the Sun side of the scan line during the austral summer at high southern latitudes viewed under large zenith angles during nominal nighttime. This flag identifies configurations in which the problem may potentially occur. A plot of the average BT differences for NOAA 9, week 41 in 1986, for zenith angles  $> 50^{\circ}$  (Figure 13) shows distinct regimes near the equator, from the equator to  $45^{\circ}$  latitude, and from  $45^{\circ}$  toward the poles.



**Figure 13.**  $T_4 - T_5$  brightness temperature differences for NOAA 9 week 41 1986. For this analysis, retrievals were restricted to zenith angle ranges  $> 50^\circ$ . The resulting fields are then zonally averaged and plotted as a function of latitude.

Generally, the channel difference increases as a function of both atmospheric moisture and satellite zenith angle as a result of the differing spectral response between the two channels and the effective increase in atmospheric pathlength at increasing satellite viewing angles. For the drier atmospheres near the poles, the  $T_{45}$  difference is influenced in a different manner than at the equator. At high zenith angle the differences are due to effects of surface emissivity, resulting in and an increased proportion of surface radiance reaching the satellite. Peaks in the high zenith angle  $T_{45}$  difference near the poles are due to sunlight leaking into the sensor (stray light). These peaks are most pronounced in the middle to high latitudes in the Southern Hemisphere. For that reason, in the Northern Hemisphere this flag is always set to 0 (pass). The latitude in the Southern Hemisphere at which the stray sunlight becomes a potential problem is a function of season. During the austral summer this problem may potentially reach the midlatitudes, whereas in austral winter, it is confined to very high latitudes. For speed of processing we have disregarded the seasonality of the latitude dependence, which may result in good pixels in the Southern Hemisphere being erroneously flagged as failing this test. As this test is later used to define overall quality levels (see below), good midlatitude Southern Hemisphere pixels at high scan angles have the potential of being assigned to the lowest quality level during austral winter.

**6.4.9. Edge test.** Pixels must not be on the first or last scan lines of a piece or on the first or last pixels in a scan line. Pixels along edges are not surrounded by other pixels and therefore are not possible to test for many of the conditions listed above that examine information in a  $3 \times 3$  pixel box.

**6.4.10. Glint test.** Glint index must be  $< 0.005 \text{ sr}^{-1}$ . The glint index is computed using the *Cox and Munk* [1954] formulation, assuming a nominal surface wind speed of  $6 \text{ m s}^{-1}$ . A value  $> 0.005 \text{ sr}^{-1}$  generally indicates significant presence of Sun glint.

## 6.5. Overall Quality Levels of Global SST Fields

The outcomes of the individual quality condition tests are subsequently combined into an overall quality level for each pixel. There are eight possible overall quality levels (levels 0–7) to which a pixel may be assigned. A quality level of 0

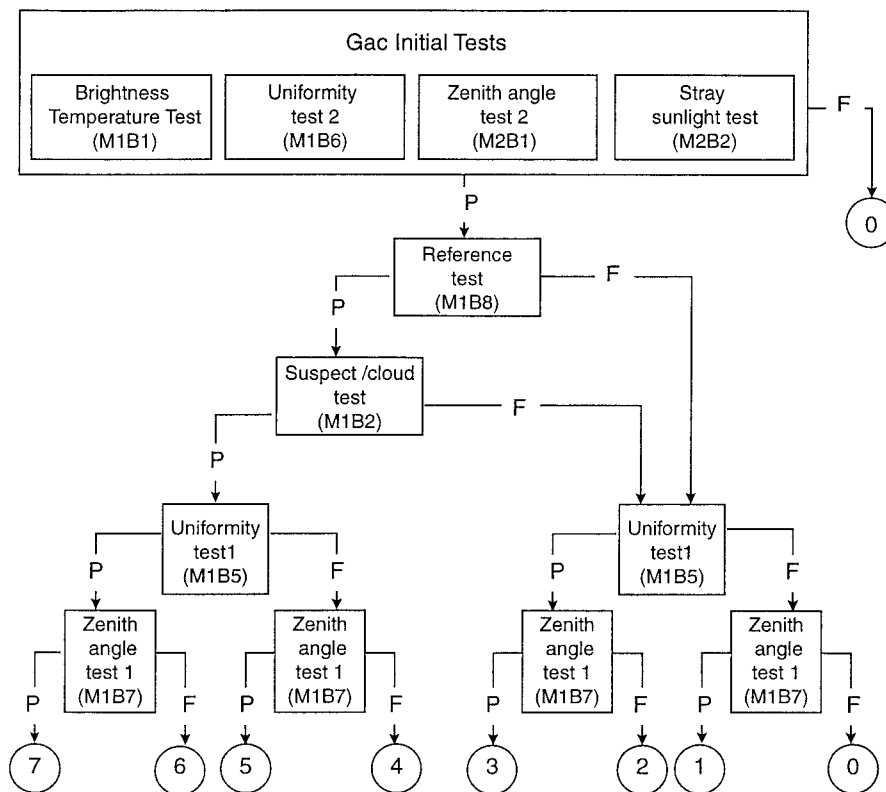
indicates very bad SST data, while level 7 is the highest quality. Overall quality levels are only determined in the global SST fields; they are not explicitly determined in the matchup database.

Pixels of the poorest quality (level 0) are identified through a few initial condition tests associated with potential gross SST errors. A pixel is automatically assigned to the lowest quality level (0) if a pixel failed any of the following four quality condition tests: (1) BT test, (2) uniformity test 2, (3) zenith angle test 2, or (4) stray sunlight test.

The seven remaining possible quality levels are assigned by evaluating various combinations of the quality condition tests. These combinations are illustrated in Figure 14. We stress that overall quality levels are provided only as guidance to users and that they are not associated with any specific error levels in SST estimates. Further, the quality scale is arbitrary, and it does not involve any proportionality (e.g., pixels with quality level 4 are not twice as good as those with quality level 2). To provide data users with an understanding of the relationship between quality levels in regard to both geographic coverage and a rough accuracy of the retrievals, Figure 15 presents a histogram of the median of differences between the retrievals for a daily nighttime 9 km resolution PFSST minus the corresponding weekly  $1^\circ$  resolution Reynolds/NCEP SST. The inset of Figure 15 gives the percentage of ocean pixels assigned to each quality level in the global field. We stress that the use in this comparison of the weekly Reynolds/NCEP SST, with its own inherent uncertainties and larger time-space resolution, produces a larger cold bias than is seen in comparison to the in situ buoy data from the PFMDB. The smaller bias seen in the PFMDB is partially due to the tight time-space resolution of the in situ buoy and satellite retrievals. Furthermore, the Reynolds/NCEP reference combines both daytime and nighttime satellite data and may result in a warmer reference field when comparing to nighttime-only retrievals. A full discussion of uncertainties in the accuracy of the weekly Reynolds/NCEP SST is beyond the scope of this paper. Nevertheless, this comparison to the Reynolds/NCEP SST provides an understanding of the accuracy trade-off between quality levels and global coverage that is not possible to determine with the limited geographic coverage of in situ data alone.

Once overall quality levels are defined for all pixels in a processing piece, the next step is to combine these values into a bin quality level. This step actually takes place during the spatial binning stage, described in detail in section 6.6. For the sake of conceptual continuity, however, we discuss here how the quality level is set for a bin. During spatial or temporal binning each pixel is assigned to a bin in the Pathfinder grid. More than one pixel may be assigned to the same bin. Which pixels are included in the bin, however, is a function of the overall quality levels for all candidate pixels. Only pixels of the highest available quality are aggregated into a bin value; pixels of lower quality are not included during binning. This is best illustrated with an example. Suppose three pixels could be assigned to bin  $N$ ; two of these pixels have a quality of 3, and the remaining pixel has a (higher) quality level 5. In this case, only the pixel with quality 5 is binned, and the two quality 3 pixels are discarded. That is, the spatial or temporal binning procedure considers only the “best” data available for a given bin.

Users of binned data may select what SST quality levels they may wish to consider in their specific application. For instance, if quantitative analyses are being performed on SST values



**Figure 14.** Combination of individual quality tests to derive an overall pixel quality level. Final quality levels are indicated in the circles. The location of the test result in the appropriate mask variable is indicated (in parenthesis) as “MXBY,” where  $X$  is 1 or 2, indicating whether the test result is in mask 1 or mask 2, and  $Y$  is the bit (1–8) in the corresponding mask variable.

(e.g., for climate studies), users will probably want to use only the best quality SST estimates. On the other hand, if the goal is to monitor patterns (e.g., frontal features), users may be willing to accept lower-quality levels, trading off SST quality for more complete geographic coverage.

#### 6.6. Spatial Binning and the Equal Area Grid

The next step in processing involves spatially binning all pixels in a piece into an equal area grid. Complete descriptions of both the spatial and temporal binning procedures used in Pathfinder processing are given by *Campbell et al.* [1995]. During this step each pixel in a processing piece is assigned to a bin in an equal area grid. The spatial grid adopted for AVHRR Oceans Pathfinder processing is based on the equal area grid adopted by the International Satellite Cloud Climatology Project. The Pathfinder grid as described by *Campbell et al.* [1995] is an integerized sinusoidal equal area grid and consists of rectangular bins or tiles arranged in zonal rows. These bins are approximately 9.28 km per side at the equator. This results in 2160 zonal rows of tiles from pole to pole (1080 in each hemisphere). There are a total of 5,940,422 bins in the global grid. The bins are arranged in rows beginning at 180° longitude and circumscribing the Earth eastward at a given latitude. The number of tiles per row varies with latitude. The rows immediately above and below the equator have 4320 bins; at the poles the number of tiles is always three. Between these two extremes the number of bins in a zonal row decreases as an approximate cosine function of latitude.

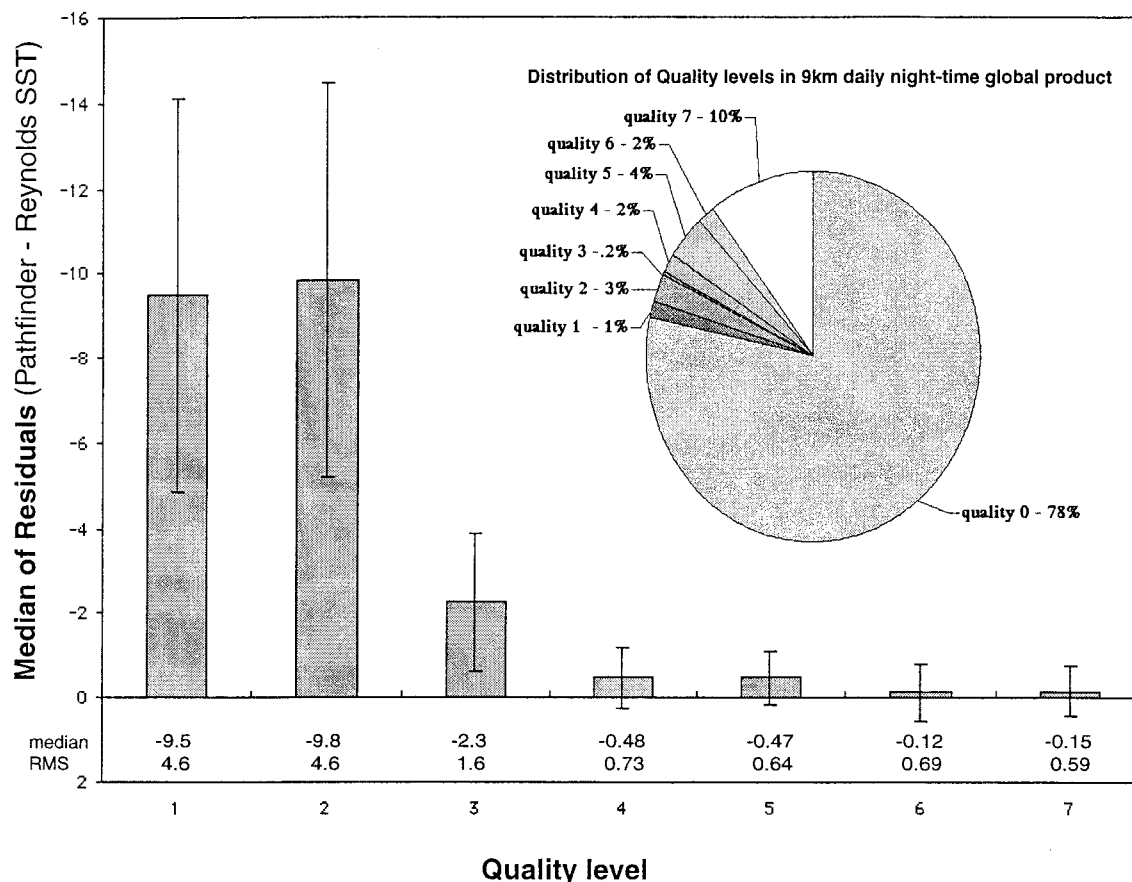
As described above, only pixels of the highest and equal

quality are assigned to a given bin. The binning procedure therefore outputs only the best data available for a given bin number. The status of the quality masks at the bin level is determined by evaluating the outcome of the quality condition tests stored in the individual pixel level masks for each of the summed pixels. For example, if any of the summed pixels failed a given quality condition test, then the equivalent test in the bin level mask is also set to failed. For each bin the bin number, quality masks, quality level, summed SST, summed squared SST, and counts are stored. The average SST value retrieved for a given bin can then be calculated from the count and the sum of the SST and represents the average of all pixels accumulated.

#### 6.7. Time Binning and the Accumulation of Daily Global Fields

The basic products generated by the NOAA/NASA AVHRR Oceans Pathfinder project are global daily fields of SST. To construct these fields, a consistent definition of a data day was adopted, as described in considerable detail by *Podestá* [1995]. The definition of the data day encompasses both a spatial and a temporal component. If either a temporal (e.g., a 24 hour period) or spatial definition (e.g., full global coverage) were to be used alone, the resulting daily product would have large gaps or discontinuities at the time or space boundary because of satellite orbital mechanics. For the Pathfinder project the spatial component defines the start and end time of a given data day and is defined as the time at which the satellite orbit track crosses the 180° meridian nearest to the equator.





**Figure 15.** Histogram of residuals for each quality level in a Pathfinder 9 km global daily nighttime SST product. Data shown are from NOAA 14 April 11, 2000. The median value of the residuals for each quality level were calculated from a 9 km resolution GAC PFSST minus the corresponding 1° weekly Reynolds SST. Error bars show the rms of the difference. The inset is a pie chart showing the percentage of total ocean pixel count in each quality level for the image. Note the use of the weekly 1° resolution Reynolds/NCEP SST in this comparison produces a larger cold bias in the resulting residuals than is seen in the comparison to the in situ buoy data from the PFMDB.

The start and end of the data day therefore changes as a function of the satellite orbit. This strict spatial definition results in large temporal discontinuities at the meridian boundary as the orbit precesses. To eliminate these discontinuities, the temporal component includes any data taken within 2 hours before or 2 hours after the meridian crossing. A data day therefore may represent data taken over a 24–28 hour period. For the Pathfinder project, ascending (daylight) and descending (nighttime) passes are processed separately. Users should be aware that the satellite crosses the equator at the 180° meridian on the descending pass ~12 hours prior to the ascending pass. Therefore, for a given data day, there are only 12 hours of temporal overlap between the global ascending and descending fields.

During the daily time binning of overlapping passes, as in the spatial binning, only the best bins of equal quality level are summed, and the equivalent quality mask is set. These interim daily fields are stored for later reevaluation of the quality level using a 3 week Pathfinder reference. The 3 week Pathfinder reference is created by the temporal binning of seven global daily files containing only quality level 7 retrievals into a week and then binning the three weekly files to create a 3 week average. Any empty bins in the 3 week Pathfinder reference are then filled using a smoothed distance-weighted interpola-

tion of the surrounding bins. A new 9 km resolution 3 week Pathfinder reference file is produced for each week. A given 3 week Pathfinder reference file is then used to reevaluate the quality level of the interim daily fields associated with the middle week of the reference file. The final quality level for each of the global daily bins is determined by comparing the interim daily field to the 3 week reference. If the difference between the daily and the 3 week Pathfinder reference is  $> \pm 2^\circ\text{C}$ , then the bin's quality is demoted one quality level. The comparison to the 3 week Pathfinder reference identifies suspect pixels that may not have been flagged by the main core of quality condition tests. For example, while the  $3 \times 3$  pixel homogeneity test eliminates pixels contaminated by nearby clouds, it is possible to have a uniform cloud covering the entire  $3 \times 3$  box, which would not be detected by a homogeneity test. The comparison to the 3 week Pathfinder reference also may identify problem pixels associated with transient atmospheric events such as absorbing aerosol and dust contamination that are currently undetectable by other methods.

The quality-controlled daily fields are delivered to NASA's Jet Propulsion Laboratory Physical Oceanography Distributive Active Archive Center (JPL PO.DAAC) for public distribution. The NOAA/NASA AVHRR Oceans Pathfinder SST data fields and PFMDB are distributed by the JPL PO.DAAC in

several spatial (9, 18, and 54 km) and temporal resolutions (daily, weekly, and monthly). Data are made available via ftp and through subsetting routines. (A description of the SST products and formats available through the JPL PO.DAAC can be found at <http://podaac-www.jpl.nasa.gov/sst/>).

## 7. Discussion

The main challenge in developing a Pathfinder global SST algorithm is to achieve relatively uniform performance throughout a wide range of atmospheric and oceanic conditions. In the past few years several aspects of the Pathfinder processing have undergone major improvements toward this end. In the first version of the processing protocol (version 1.0) the PFSST algorithm was based on a modified NLSST algorithm [Walton *et al.*, 1998a] for which separate coefficients were derived for two distinct water vapor regimes. A time-dependent term was then added to the PFSST algorithm (version 2.0) to account for observed temporal trends in the SST residuals (the difference between observed and estimated SST values). This term was replaced in version 3.0 by the estimation of algorithm's coefficients on a monthly basis, which produced more consistent results. In the current major revision (version 4.0) of the PFSST algorithm a robust regression procedure was introduced during coefficient determination. As part of the data quality assessment component of the processing protocol, decision trees were developed to identify objectively cloud contaminated/suspect retrievals. A larger number of pixel quality levels were added to the PFSST products to provide more flexibility to users who needed to make trade-offs between product quality and data coverage.

Comparisons with in situ buoys indicate that the global accuracy of the current Pathfinder algorithm is  $0.02^\circ \pm 0.5^\circ\text{C}$ , while comparison with a radiometric reference of skin temperature (MAERI) yields  $0.14^\circ \pm 0.31^\circ\text{C}$ . This Pathfinder—MAERI difference is comparable to the difference between the MAERI and other bulk SST measurements, indicating that in situ buoy calibration and skin-bulk temperature differences are not an issue in the PFSST. These global statistics, however, can be misleading. As Barton [1995] pointed out, SST algorithms assume a mean state of the atmosphere (e.g., a typical shape of water vapor and temperature profiles). A similar statement can be made about typical oceanic conditions (e.g., no upper level stratification assumed when comparing to in situ SST measurements). When conditions deviate from the mean atmosphere and ocean conditions, errors arise in SST retrievals. Deviations from first-guess conditions are more likely in a global algorithm than in regionally tuned algorithms, and this should be kept in mind when evaluating global SST estimates. Furthermore, in the case of statistically derived global SST algorithms, the first-guess conditions will be the average of conditions at all the matchup locations and times used in coefficient estimation. We stress that this average will be weighted by the relative distribution of matchups, which has and will continue to change in time. The performance of an SST algorithm for a given set of atmospheric and oceanic conditions therefore depends not only on how close those conditions are to the average state but also on how well represented those conditions are in the matchup set used to derive the algorithm coefficients.

**Acknowledgments.** We would like to thank V. Halliwell, J. Splain, and S. Walsh for software and data processing support. We thank Dr.

S. Shenoi for providing the water vapor analysis and the comments of two anonymous reviewers. This work was supported by the NOAA/NASA Pathfinder program (NAG56326).

## References

- Anding, D., and R. Kauth, Estimation of sea surface temperature from space, *Remote Sens. Environ.*, **1**, 217–220, 1970.
- Barton, I. J., Satellite-derived sea surface temperatures: A comparison between operational, theoretical and experimental algorithms, *J. Appl. Meteorol.*, **31**, 433–442, 1992.
- Barton, I. J., Satellite-derived sea surface temperatures: Current status, *J. Geophys. Res.*, **100**, 8777–8790, 1995.
- Barton, I. J., and R. P. Checcet, Comparison and optimization of AVHRR sea surface temperature algorithms, *J. Atmos. Oceanic Technol.*, **6**, 1083–1089, 1989.
- Barton, I. J., A. M. Zavody, D. M. O'Brien, D. R. Cutten, R. W. Saunders, and D. T. Llewellyn-Jones, Theoretical algorithms for satellite-derived sea surface temperatures, *J. Geophys. Res.*, **94**, 3365–3375, 1989.
- Brown, J., O. Brown, and R. Evans, Calibration of advanced very high resolution radiometer Infrared Channels: A new approach to non-linear correction, *J. Geophys. Res.*, **98**, 18,157–18,269, 1993.
- Brown, O. B., J. Brown, and R. Evans, Calibration of AVHRR infrared observations, *J. Geophys. Res.*, **90**, 11,667–11,677, 1985.
- Campbell, J., J. M. Blaisdell, and M. Darzi, Spatial and temporal binning algorithms, *NASA Tech. Memo.*, **104566**, 32, Appendix A, pp. 63–65, 1995.
- Cayula, J. F., and P. Cornillon, Cloud detection from a sequence of SST images, *Remote Sens. Environ.*, **55**, 80–88, 1996.
- Clark, L. A., and D. Pregibon, Tree-based models, in *Statistical Models in S*, edited by J. Chambers and T. H. Hastie, pp. 377–419, Wadsworth, Belmont, Calif., 1992.
- Cleveland, W. S., *Visualizing Data*, 360 pp., Hobart, Summit, N. J., 1993.
- Cornillon, P., C. Gilman, C. L. Stramma, O. Brown, R. Evans, and J. Brown, Processing and analysis of large volumes of satellite-derived thermal infrared data, *J. Geophys. Res.*, **92**, 12,993–13,002, 1987.
- Cox, C., and W. Munk, Measurements of the roughness of the sea surface from photographs of the Sun's glitter, *J. Opt. Soc. Am.*, **44**, 838–850, 1954.
- Derrien, M., B. Farki, L. Harang, H. LeGléau, A. Noyalet, D. Pochic, and A. Sairouni, Automatic cloud detection applied to NOAA-11/AVHRR imagery, *Remote Sens. Environ.*, **46**, 246–267, 1993.
- Kearns, E. J., J. A. Hanafin, R. H. Evans, P. J. Minnett, and O. B. Brown, An independent assessment of Pathfinder AVHRR sea surface temperature accuracy using the Marine—Atmosphere Emitted Radiance Interferometer, *Bull. Meteorol. Soc.*, **81**, 1525–1536, 2000.
- Lanzante, J. R., Resistant, robust and non-parametric techniques for the analysis of climate data: Theory and examples, including applications to historical radiosonde station data, *Int. J. Climatol.*, **16**, 1197–1226, 1996.
- Luo, G., P. A. Davis, L. B. Stowe, and E. P. McClain, A pixel-scale algorithm of cloud type, layer and amount for AVHRR data, part I, Nighttime, *J. Atmos. Oceanic Technol.*, **12**, 1013–1037, 1995.
- McClain, E. P., W. G. Pichel, and C. C. Walton, Comparative performance of AVHRR-based multichannel sea surface temperatures, *J. Geophys. Res.*, **90**, 11,587–11,601, 1985.
- Minnett, P. J., The regional optimization of infrared measurements of sea surface temperature from space, *J. Geophys. Res.*, **95**, 13,497–13,510, 1990.
- Minnett, P. J., Consequences of sea surface temperature variability on the validation and application of satellite measurements, *J. Geophys. Res.*, **96**, 18,475–18,489, 1991.
- Podestá, G. P., Case studies for SeaWiFS calibration and validation, part 3, SeaWiFS global fields: What's in a day?, *NASA Tech. Memo.*, **10456**, 27 pp., 1995.
- Prabhakara, C., G. Dalu, and V. G. Kunde, Estimation of sea surface temperature from remote sensing in the 11 to 13-um window region, *J. Geophys. Res.*, **79**, 5039–5044, 1974.
- Rao, C. R. N., and J. Chen, Inter-satellite calibration linkages for the visible and near-infrared channels of the advanced very high resolution radiometers on the NOAA-7, -9, and -11 spacecraft, *Int. J. Remote Sens.*, **16**, 1931–1942, 1995.
- Rao, C. R. N., and J. Chen, Post launch calibration of the visible and near-infrared channels of the advanced very high resolution radiom-

- eters on the NOAA-14 spacecraft, *Int. J. Remote Sens.*, **17**, 2743–2747, 1996.
- Reynolds, R. W., and T. M. Smith, Improved global sea surface temperature analysis using optimum interpolation, *J. Clim.*, **7**, 929–948, 1994.
- Rousseeuw, P. J., and A. M. Leroy, Robust regression and outlier detection, 329 pp., John Wiley, New York, 1987.
- Saunders, R. W., and K. T. Kriebel, An improved method for detecting clear sky and cloudy radiances from AVHRR data, *Int. J. Remote Sens.*, **9**, 123–150, 1988.
- Schluessel, P., W. J. Emery, H. Grassl, and T. Mammen, On the skin—bulk temperature difference and its impact on satellite remote sensing of sea surface temperature, *J. Geophys. Res.*, **95**, 13,341–13,356, 1990.
- Schwalb, A., The TIROS-N/NOAA A-G satellite series, *NOAA Tech. Memo., NESS 95*, 75 pp., 1978.
- Venables, W. N., and B. D. Ripley, Tree-based methods, in *Modern Applied Statistics With S Plus*, 3rd ed., chap. 10, pp. 329–345, Springer-Verlag, New York, 1999.
- Walton, C. C., Nonlinear multichannel algorithm for estimating sea surface temperature with AVHRR satellite data, *J. Appl. Meteorol.*, **27**, 115–124, 1988.
- Walton, C. C., E. P. McClain, and J. F. Sapper, Recent changes in satellite based multichannel sea surface temperature algorithms, paper presented at Marine Technology Society Meeting MTS' 90, Washington, D. C., September 1990.
- Walton, C. C., W. G. Pichel, and J. F. Sapper, The development and operational application of non-linear algorithms for the measurement of sea surface temperatures with the NOAA polar-orbiting environmental satellites, *J. Geophys. Res.*, **103**, 27,999–28,012, 1998a.
- Walton, C. C., J. W. Sullivan, C. R. N. Rao, and M. P. Weinreb, Correction for detector nonlinearities and calibration inconsistencies of the infrared channels of the advanced very high resolution radiometer, *J. Geophys. Res.*, **103**, 3323–3337, 1998b.
- Weinreb, M. P., G. Hamilton, S. Brown, and R. J. Koczor, Nonlinearity corrections in calibration of advanced very high resolution radiometer infrared channels, *J. Geophys. Res.*, **95**, 7381–7388, 1990.
- Zavody, A. M., J. T. Mutlow, and D. T. Llewellyn-Jones, A radiative transfer model for sea surface temperature retrieval for the along-track scanning radiometer, *J. Geophys. Res.*, **100**, 937–952, 1995.

---

R. Evans, K. A. Kilpatrick, and G. P. Podestá, Rosenstiel School of Marine and Atmospheric Science, University of Miami, 4600 Rickenbacker Cswy., Miami, FL 33149–1098. (kkilpatrick@rsmas.miami.edu)

(Received September 8, 1999; revised December 20, 2000; accepted January 25, 2001.)

