

Toward Improved Validation of Satellite Sea Surface Skin Temperature Measurements for Climate Research

C. J. DONLON,* P. J. MINNETT,⁺ C. GENTEMANN,[#] T. J. NIGHTINGALE,[@] I. J. BARTON,& B. WARD,** AND M. J. MURRAY⁺⁺

*CEC-Joint Research Centre, Institute for Environment and Sustainability, Inland and Marine Waters Unit, Ispra, Italy

⁺Meteorology and Physical Oceanography Division, Rosenstiel School of Marine and Atmospheric Science, University of Miami, Miami, Florida

[#]Remote Sensing Systems, Santa Rosa, California

[@]Space Science and Technology Department, Rutherford Appleton Laboratory, Chilton, Didcot, Oxon, United Kingdom

[&]CSIRO Marine Research, Hobart, Tasmania, Australia

**Cooperative Institute for Marine and Atmospheric Studies, Rosenstiel School of Marine and Atmospheric Science, University of Miami, Miami, Florida

⁺⁺Space Science and Technology Department, Rutherford Appleton Laboratory, Chilton, Didcot, Oxon, United Kingdom

(Manuscript received 6 April 2001, in final form 27 July 2001)

ABSTRACT

A poor validation strategy will compromise the quality of satellite-derived sea surface temperature (SST) products because confidence limits cannot be quantified. This paper addresses the question of how to provide the best operational strategy to validate satellite-derived skin sea surface temperature (SST_{skin}) measurements. High quality in situ observations obtained using different state-of-the-art infrared radiometer systems are used to characterize the relationship between the SST_{skin} , the subsurface SST at depth (SST_{depth}), and the surface wind speed. Data are presented for different oceans and seasons. These data indicate that above a wind speed of approximately 6 m s^{-1} the relationship between the SST_{skin} and SST_{depth} is well characterized for both day- and nighttime conditions by a cool bias of $-0.17 \pm 0.07 \text{ K}$ rms. At lower wind speeds, stratification of the upper-ocean layers during the day may complicate the relationship, while at night a cooler skin is normally observed. Based on these observations, a long-term global satellite SST_{skin} validation strategy is proposed. Emphasis is placed on the use of autonomous, ship-of-opportunity radiometer systems for areas characterized by prevailing low-wind speed conditions. For areas characterized by higher wind speed regimes, well-calibrated, quality-controlled, ship and buoy SST_{depth} observations, corrected for a cool skin bias, should also be used. It is foreseen that SST_{depth} data will provide the majority of in situ validation data required for operational satellite SST validation. We test the strategy using SST_{skin} observations from the Along Track Scanning Radiometer, which are shown to be accurate to approximately 0.2 K in the tropical Pacific Ocean, and using measurements from the Advanced Very High Resolution Radiometer. We note that this strategy provides for robust retrospective calibration and validation of satellite SST data and a means to compare and compile in a meaningful and consistent fashion similar datasets. A better understanding of the spatial and temporal variability of thermal stratification of the upper-ocean layers during low-wind speed conditions is fundamental to improvements in SST validation and development of multisensor satellite SST products.

1. Introduction

It has long been realized that global ocean measurements of sea surface temperature (SST) collected by radiometers on satellites have made a major contribution to climate research. In addition to its role as a "global thermometer," the SST is important in coupling the ocean and atmosphere through exchanges of heat, momentum, moisture, and gases. SST products are required by operational ocean analysis and prediction systems to

properly constrain the upper-ocean circulation and thermal structure and SST is a key parameter in other oceanographic fields such as coastal oceanography and monitoring of biological resources. The high spatial and radiometric resolution, the regular sampling, and the synoptic perspective of spaceborne sensors make them well suited to improving SST measurement capability. Technological advances and innovative design have resulted in new generations of satellite instruments [e.g., the Along Track Scanning Radiometer (ATSR; Mutlow et al. 1994)], the Tropical Rainfall Measuring Mission (TRMM) Microwave Imager (Kummerow et al. 1998; Wentz et al. 2000) and, the Moderate Resolution Imaging Spectroradiometer (MODIS; Esaias et al. 1998) having higher accuracy and capability compared to their

Corresponding author address: Dr. Craig Donlon, CEC-Joint Research Centre, Institute for Environment and Sustainability, Inland and Marine Waters Unit, Ispra, I-21020 TP272, Italy.
E-mail: craig.donlon@jrc.it

predecessors. This steady improvement has now reached the stage where detailed investigation of the processes and feedback mechanisms that govern the spatial and temporal dynamics of the lower atmosphere and upper ocean at scales from a few kilometers to the quasi-synoptic global is possible using satellite observations.

A scientific priority for earth observation over the next decade is to seek the best combination of tools (satellite observations, in situ measurements, and numerical models) to optimize the information content of these data by promoting interdisciplinary and multisensor research. While long-term regional and global satellite SST products do exist (e.g., Vasquez et al. 1998; Kilpatrick et al. 2001), they are often limited in terms of spatial resolution and consistency due to uneven sensor performance. A new generation of satellite SST products is required (e.g., Smith 2000) building on the complementary aspects of different satellite instruments by merging data to provide both increased fidelity, consistency, and resolution. In fact the Global Ocean Data Assimilation Experiment, (Le Traon et al. 1999) has recently initiated a pilot project to develop high spatial and temporal resolution SST data products (Donlon 2001). Against this background, it is important that a robust and pragmatic approach to the continuing validation of global satellite-derived SST products is implemented to assure accurate and dependable SST data products.

In the past, accurate in situ radiometers mounted on research vessels have obtained SST measurements contemporaneous with satellite measurements that are used to validate the atmospheric correction strategies necessary to derive SST from top of the atmosphere radiance measurements. Collectively, these provide robust and valuable validation data (e.g., Kearns et al. 2000) but, due to the often limited geographic and temporal coverage of individual campaigns, the cost and difficulty of maintaining instrumentation, this approach falls short of a continuing global satellite SST validation strategy. If in situ radiometers remain the only SST validation data source, it is clear that the SST validation dataset will always be limited, both in spatial distribution and quantity.

This paper specifically addresses the need for continuous global validation of infrared satellite sea surface temperature measurements. Section 2 presents state-of-the-art in situ datasets that characterize the relationship between the radiometrically measured SST and the subsurface temperature at depth. Section 3 outlines a strategy for the operational use of existing oceanographic infrastructure and highlights problems and the need for new autonomous instrumentation in certain conditions. Section 4 applies the methods discussed in sections 3 and 4 to satellite data. Finally, section 5 discusses a global validation SST strategy and presents our final conclusions and recommendations.

2. Measurements of sea surface temperature

The vertical temperature structure of the upper ocean (~ 10 m) is both complex and variable depending on the level of shear-driven ocean turbulence and the air-sea fluxes of heat, moisture, and momentum. Thus, every SST observation depends on the measurement technique and sensor that is used, the vertical position of the measurement within the water column, the local history of all component heat flux conditions, and the time of day the measurement was obtained. The vertical structure of SST can be generally classified as follows.

- The interface SST, SST_{int} , is the temperature of an infinitely thin layer at the exact air-sea interface. It represents the temperature at the top of the SST_{skin} layer (and hence the top of the temperature gradient in that layer) and cannot be measured using current technology.
- The skin SST, SST_{skin} , is a temperature measured by a radiometer at depth within a thin layer ($\sim 500 \mu\text{m}$) at the water side of the air-sea interface where conductive and diffusive heat transfer processes dominate. A strong temperature gradient is characteristically maintained in this thin layer sustained by the magnitude and direction of the ocean-atmosphere heat flux. Thus, SST_{skin} varies according to depth within the layer and because the penetration depth of the emitted radiation is a function of the wavelength of the radiation, the value of SST_{skin} varies dependent on the wavelength used for the measurement. This is the basis of measuring the skin temperature gradient using infrared interferometry at wavelengths shorter than $5 \mu\text{m}$ (McKeown et al. 1995). Consequently, SST_{skin} should always be quoted at a specific wavelength in the water column, for example, $SST_{skin 10.5 \mu\text{m}}$. However, over the parts of the infrared spectrum used to make the measurements of SST_{skin} presented in this study, the variation in the penetration depth is very small, and the values of SST_{skin} measured by ideal infrared radiometers are expected to vary by less than 0.01 K, and the wavelength dependence of SST_{skin} can be ignored.
- The subskin SST, $SST_{subskin}$, is representative of the SST at the bottom of the SST_{skin} temperature gradient (layer) where molecular and viscous heat transfer processes begin to dominate. It varies on a timescale of minutes and may be influenced by solar warming. $SST_{subskin}$ may be approximated by measurements made by a low-frequency (6–10 GHz) microwave radiometer. In this region of the electromagnetic spectrum, the penetration depth in seawater is much greater, resulting in measurements depths of greater than 1 mm.
- The subsurface SST, SST_{depth} , (traditionally referred to as a “bulk” SST) considers any temperature within the water column beneath the $SST_{subskin}$ where turbulent heat transfer processes dominate. It may be significantly influenced by local solar heating and has a time-

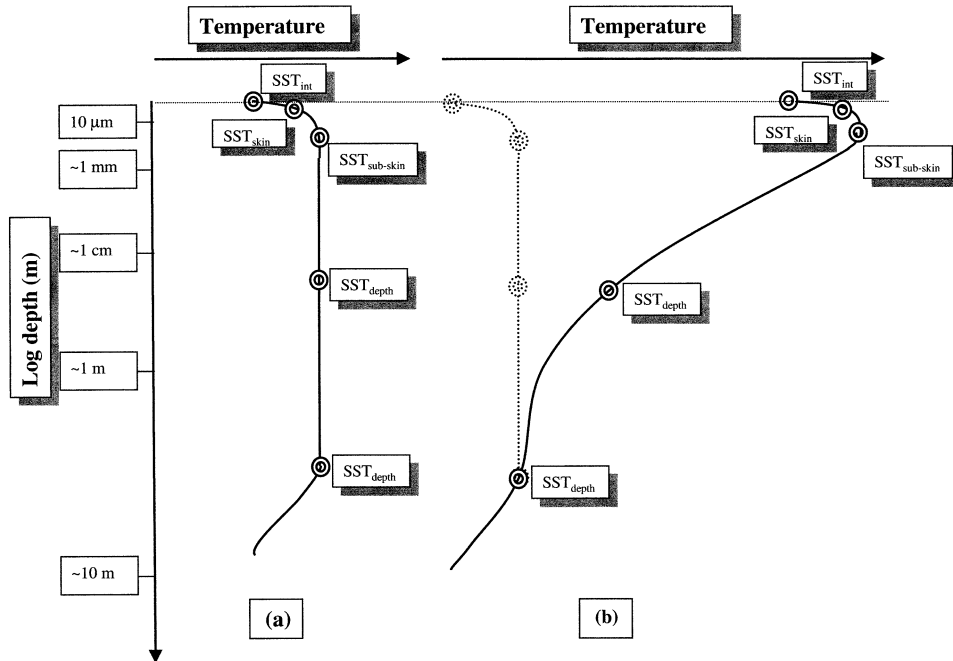


FIG. 1. Idealized temperature profiles of the near-surface layer (~ 10 -m depth) of the ocean during (a) nighttime and daytime during strong wind conditions and (b) daytime low-wind speed conditions and high insolation resulting thermal stratification of the surface layers.

scale of hours and typically varies with depth. Consequently, SST_{depth} should always be quoted at a specific depth in the water column; for example, $SST_{5\text{ m}}$ refers to the SST at a depth of 5 m. SST_{depth} is measured using traditional temperature sensors mounted on buoys, profilers, and ships at any depth beneath SST_{subskin} .

There is considerable observational evidence confirming these relationships (e.g., Saunders 1967; Hasse 1971; Grassl 1976; Katsaros 1977; Paulson and Simpson 1981; Schlüssel et al. 1990; Fairall et al. 1996; Kent et al. 1996; Donlon and Robinson 1997; Minnett and Hanafin 1998; Donlon et al. 1999a; Minnett and Ward 2000; Ward and Minnett 2002). Figure 1 illustrates schematically the importance of referencing the mean depth or wavelength at which an SST measurement is determined when considering upper-ocean SST. Figure 1a shows the characteristic thermal structure at night or during moderate to strong winds during the day that homogenize the temperature in the upper-water layers. SST_{subskin} is similar to SST_{depth} at all depths but is characteristically warmer than the cooler SST_{skin} . Figure 1b depicts the characteristic situation for late morning–early afternoon following a period of light or absent wind and insolation. Thermal stratification of the upper-ocean layers has occurred resulting in significant temperature differences between SST_{int} , SST_{skin} , SST_{subskin} , and SST_{depth} . This underscores the motivation to refine the traditional reference to bulk SST, universally used in oceanography and meteorology, to the more exact parameter, SST_{depth} . We note that the cloud-free conditions described by Fig. 1b are

favorable for infrared remote sensing of SST from satellite instruments and in such cases, the significant vertical variation of SST demands careful attention in terms of SST data product conception, validation, and interpretation. In the future, if microwave, infrared, and in situ SST datasets are used together (e.g., Donlon 2001), understanding and reconciling these differences is exceptionally important.

In order to quantify vertical thermal differences the parameter ΔT is traditionally used, which is defined as

$$\Delta T_{\text{subskin}} = SST_{\text{skin}} - SST_{\text{subskin}}. \quad (1a)$$

A negative value of ΔT computed using (1a) corresponds to a cool skin temperature. (Note that some authors use $SST_{\text{subskin}} - SST_{\text{skin}}$ that results in a positive ΔT value for a cool skin temperature.) SST_{subskin} is extremely difficult to measure requiring accurate multi-spectral infrared instruments and has only been proven in the laboratory (e.g., McKeown and Asher 1997; McKeown et al. 1995). Measurements using microwave radiometers provide SST_{subskin} and may in the near future attain the necessary accuracy for direct computation of ΔT from satellite data alone (Wentz and Meissner 1999). However, in situ microwave instruments have not been deployed widely and consequently (1a) cannot be readily applied. Instead, SST_{subskin} is typically replaced by SST_{depth} , which can be measured using traditional contact temperature sensors so that (1a) becomes

$$\Delta T_{\text{depth}} = SST_{\text{skin}} - SST_{\text{depth}}. \quad (1b)$$

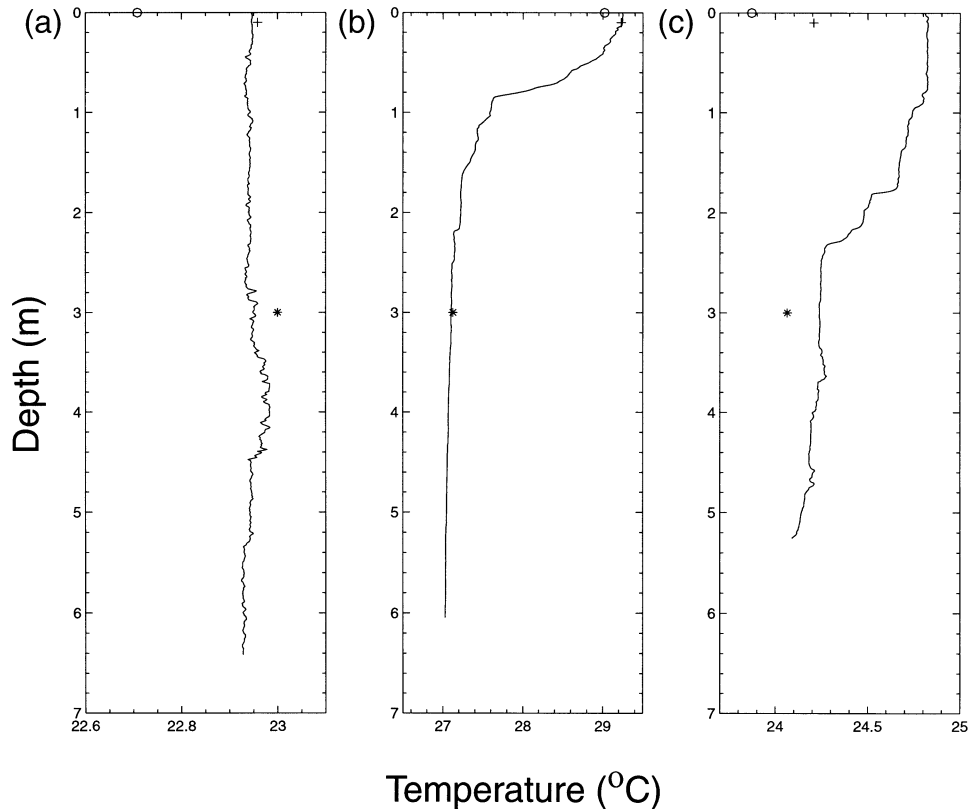


FIG. 2. Daytime and nighttime vertical temperature profiles at the near-ocean surface in the Gulf of California obtained using the SkinDeEP profiling instrument. In each plot, the circle is the skin temperature measured by M-AERI, the plus signs the in situ temperature measured at a depth of about 5 cm, and the star is the temperature measurement from a thermosalinograph on the ship at a nominal depth of 3 m. (a) SkinDeEP profiles at 0655 LT, (b) SkinDeEP profile at 1309 LT, and (c) SkinDeEP profile at 1920 LT.

Clearly it is important to recognize that (1b) now includes both the skin temperature deviation and the effect of thermal stratification that may be present in the upper ocean so that ΔT_{depth} may be significantly different from ΔT .

a. Decoupling of SST_{skin} and SST_{depth} due to thermal stratification

Figure 2 shows three vertical profiles taken by an ascending instrument package called the Skin Depth Experimental Profiler (SkinDeEP; Ward and Minnett 2001). A complete description of SkinDeEP can be found in Ward et al. (2001, manuscript submitted to *J. Atmos. Oceanic Technol.*). Over 1000 profiles were made by SkinDeEP on the *MOCE-5* cruise in the Gulf of California in October 1999, and Fig. 2 shows just three profiles at three different times of day, on three different days (a more complete description of the data will appear elsewhere). Also shown as circles are SST_{skin} observations measured by a Marine-Atmospheric Emitted Radiance Interferometer (M-AERI, see section 3a(1) below), the $SST_{0.1\text{ m}}$ measured by a thermistor (plus sign), and the $SST_{3\text{ m}}$ recorded by a thermosalinograph (asterisk).

Figure 2a shows measurements taken just after sunrise (0655 local time) demonstrating thermal structure nearly identical to the schematic diagram shown in Fig. 1a. As the surface layer becomes progressively warmed during the day, strong stratification occurs, which is shown in Fig. 2b taken at approximately 1309 local time. This is nearly identical to Fig. 1b. Figure 2c shows profile measurements made after sunset at approximately 1720 local time that, although there was no solar heating, do not reflect the homogeneous temperature structure shown in Fig. 1a. Instead, a complex profile is observed probably due to interleaving and overturning processes causing gradients in both the horizontal and vertical directions. In all cases, the absorption of shortwave radiation in the molecular boundary skin layer is not enough to overcome the heat loss due to the sensible and latent heat fluxes, and the SST_{skin} remains cooler than the water beneath.

Considering Fig. 2, it is readily apparent the depth at which SST_{depth} is measured is of considerable importance in understanding how representative it may be of subsurface temperature: $\Delta T_{0.1\text{ m}}$ remains negative whereas $\Delta T_{3\text{ m}}$ changes sign and has considerably more variability depending on the dynamic structure of the strat-

ification. In the context of satellite SST retrievals during low wind speed, high-insolation conditions, the data presented in Fig. 2 demonstrate that the SST_{skin} and the SST_{depth} are related but quasi-independent variables that are often different requiring knowledge of the time of day and depth at which the SST_{depth} is measured for a correct interpretation. This poses significant implications for the validation, interpretation, and merging of complimentary satellite SST datasets. Further research characterizing and modeling the spatial and temporal structure of thermal stratification is vital to understanding both the accuracy of SST_{skin} , derived from the measurements of infrared radiometers on satellites, and how best to merge these data with measurements of SST_{subskin} from microwave radiometers on satellites.

b. Measurements of SST from space

Accurate and dependable instrument prelaunch characterization and postlaunch self-calibration underpin the integrity of all satellite-derived SST estimates. However, the quality and credibility of derived data products depends on accounting comprehensively for the uncertainties associated with the measurement itself, the accuracy of atmospheric correction algorithm and, the changes within satellite sensors throughout their lifetime. Thus SST product validation is distinct from sensor calibration (e.g., Mutlow et al. 1994). The fundamental measurement made by a self-calibrating satellite infrared radiometer is the upwelling radiance at the height of the satellite sensor specified for a number of spectral intervals and atmospheric pathlength. For views of the sea surface, the calibrated radiance is composed of sea surface, direct atmospheric and reflected surface radiance (due to the nonunity emissivity of the ocean surface). Atmospheric absorption and emission along the atmospheric path between the sea surface and satellite sensor (ignoring aerosol contributions) is principally by water vapor, carbon dioxide, and ozone (e.g., Saunders and Edwards 1989; Zavody et al. 1995). SST measurements are derived by compensating for unwanted radiance components reflected at the sea surface and the atmospheric attenuation of the emitted oceanic radiance. This is achieved using a weighted combination of spectral and view-specific radiant temperatures or “brightness temperatures” (i.e., the temperature of a blackbody emitter that produces that radiance, weighted by the radiometer’s relative spectral response function, for a particular spectral interval).

Considerable effort has been invested in developing semiempirical infrared atmospheric correction algorithms by regression of satellite brightness temperatures and quality-controlled in situ, subsurface SST_{depth} observations; these lead to an estimate of the subsurface temperature, referred to here as “pseudo- SST_{depth} ” ($pSST_{\text{depth}}$) algorithms (for a review see Barton 1995). An alternative, geophysically based, atmospheric correction strategy uses detailed radiative transfer calcu-

lations to predict a set of theoretical brightness temperatures for a satellite radiometer, given a set of atmospheric conditions and associated SST_{skin} values (e.g., Llewellyn-Jones et al. 1984; Minnett 1986, 1990; Zavody et al. 1995; Merchant et al. 1999) thereby specifying explicitly the effect of the atmosphere. The model-derived brightness temperatures are then used in a similar regression procedure as for $pSST_{\text{depth}}$ (i.e., matching top of the atmosphere brightness temperatures with the surface temperatures), but in this case providing a retrieval of SST_{skin} . It is important to recognize that this approach generates algorithms that assume no relation between the subsurface SST_{depth} and the SST_{skin} .

In contrast to SST_{skin} algorithms, $pSST_{\text{depth}}$ algorithms implicitly assume that the difference between the SST_{skin} and the SST_{depth} is constant or correlated with the effects of the overlying atmosphere and well specified by the in situ observations that are used to define the algorithm (typically comprising of several thousand contemporaneous satellite and in situ data points). Advanced $pSST_{\text{depth}}$ algorithms generate monthly regression coefficients after in situ data quality control procedures have been completed (e.g., Kilpatrick et al. 2001). The $pSST_{\text{depth}}$ approach may accommodate the effects of diurnal stratification by simply providing a separate daytime $pSST_{\text{depth}}$ algorithm. For example, differences between SST_{skin} and SST_{depth} during the day are significant for the Advanced Very High Resolution Radiometer (AVHRR) sensor considering the variable local overpass times, caused by orbit drift, and the range of local time across its wide swath. Such differences highlight the reason why daytime $pSST_{\text{depth}}$ algorithms may be problematic to derive but, perhaps more importantly, difficult to validate using in situ SST_{depth} observations (see Barton 1998). Nevertheless, in the following sections, we will show why the $pSST_{\text{depth}}$ approach has performed surprisingly well during the last 20 years.

3. Reconciling SST_{skin} and SST_{depth}

Donlon et al. (1999a) use extensive in situ observations obtained during independent experiments in the Atlantic Ocean to demonstrate the relationship between the surface wind speed and $\Delta T_{5\text{ m}}$. SST_{skin} was measured using a scanning infrared sea surface temperature radiometer (SISTeR). The Scanning Infrared Sea Surface Temperature Radiometer (SISTeR) is a compact and robust chopped self-calibrating filter radiometer. The instrument is divided into three compartments containing the fore optics, scan mirror, and reference blackbodies, and a small-format PC with signal processing and control electronics. The fore optics compartment contains a pyroelectric detector and preamplifier, mounted onto an assembly containing filter wheel with three narrow-band filters centered at 3.7, 10.8, and 12.0 μm , and a black rotating chopper. The entire optical system is referred to two highly accurate reference blackbodies, one floating near to ambient temperature and the other at

approximately 10 K above. Typically, SISTeR radiances are sampled every 0.8 s with the 10.8- μm filter and SST_{skin} are calculated from the upwelling ocean radiance samples, corrected for a small reflected sky radiance term with the complementary sky samples (Donlon and Nightingale 2000). $\text{SST}_{5\text{m}}$ was measured using a SeaBird thermosalinograph via a pumped seawater supply, and warming due to pumps and internal conduits accounted for using an accurate thermistor located at the water intake aperture. Wind speeds were measured using a sonic anemometer at a height of 23 m corrected for ship movements. In all cases, at wind speeds greater than 6 m s^{-1} , the variability of $\Delta T_{5\text{m}}$ was diminished and the mean value of $\Delta T_{5\text{m}}$ approximates a constant cool bias of $-0.14 \pm 0.1 \text{ K rms}$. While consistent, a weakness of this conclusion, in terms of wider applications, stems from the fact that these data were all collected in the Atlantic area using the same SISTeR radiometer.

Other shipboard radiometer systems have been deployed in different oceans and different seasons as part of satellite SST validation campaigns and air–sea interaction studies. Most notable are the deployments of the M-AERI (Minnett et al. 2001a), an accurate spectroradiometer (P. J. Minnett 2001, unpublished manuscript) in the Pacific Ocean and the DAR001 radiometer in the western Pacific (Barton 2001). In addition, new datasets have been collected using the SISTeR radiometer in the tropical Pacific. Figure 3 shows the location of observations made by the M-AERI and DAR011 radiometer systems and the SISTeR radiometer (including the data reported by Donlon et al. 1999a), that we use in the following sections to further investigate the importance of ΔT_{depth} in the context of satellite SST validation.

a. *Revelle 1997 Pacific Ocean cruise*

The M-AERI is a robust, sea-going Fourier transform infrared spectroradiometer that operates in the spectral range of approximately 3–18 μm providing spectra with a resolution of about 0.5 cm^{-1} . It uses two infrared detectors to achieve this wide spectral range cooled to approximately 78 K by a Stirling cycle mechanical cooler to reduce the noise equivalent temperature difference (NE Δ T) to $<0.1 \text{ K}$ for a target at about 270 K. The M-AERI includes two internal blackbody cavities for accurate real-time calibration. A scan mirror directs the field of view from the interferometer to each of the blackbody calibration targets or to the environment from nadir to zenith. The mirror is programmed to step through a preselected range of angles, including calibration measurements of the blackbodies before and after each set of environmental measurements. Under normal operating conditions, spectra of sea surface and atmospheric emission from which SST_{skin} can be calculated is provided every 10–12 min.

The M-AERI data used here are from the cruise of

the R/V *Roger Revelle* from Hawaii to New Zealand during September and October 1997. The ship track crossed both Northern and Southern Hemisphere trade wind zones, the equatorial region including the inter-tropical convergence zone, and the westerlies of the Southern Hemisphere. The M-AERIs were mounted high enough so that it viewed undisturbed water, ahead of the ship's bow wave, with a view angle of 45° . Prior to the cruise, the absolute calibration of both M-AERIs was checked by measuring the temperature of a blackbody target at a known temperature, and uncertainties were less than 0.03 K. Wind speed and direction were measured using an R. M. Young propeller and vane anemometer, mounted at a height of approximately 12 m, and the $\text{SST}_{5\text{m}}$ measurements were made using a SeaBird SBE21 thermosalinograph with a remote thermometer mounted close to the seawater intake.

b. *RV Franklin Western Pacific cruise*

The R/V *Franklin* obtained observations in the western Pacific Ocean along the ship track shown in Fig. 3 in March 1999. SST_{skin} was measured in undisturbed water using a DAR011 infrared radiometer mounted on the handrail and $\text{SST}_{1\text{m}}$ was measured using a thermosalinograph system. The DAR011 radiometer is a single-channel filter radiometer with a central wavelength of 10.8 μm and a bandwidth of approximately 1.0 μm . The radiometer uses a pyroelectric detector and has an NE Δ T approximately 0.05 K and two internal blackbody targets for absolute calibration. It has an objective scan mirror to cyclically sample the ocean surface (7 min), the cold blackbody (1 min), the sky radiance (1 min), and the hot blackbody (1 min). Wind speed and direction were measured using a Rimco Cup anemometer at a height of 12 m corrected for ship motion.

c. *ARM-JAMSTEC Nauru 99 cruise*

In situ data were obtained in June and July 1999 from the R/V *Mirai* during the joint U.S. Department of Energy Atmospheric Radiation Measurement and Japanese Marine Science and Technology Center Nauru 99 program in the tropical east Pacific Ocean. Figure 3 shows the ship track of the *Mirai*. SST_{skin} were recorded with a SISTeR radiometer mounted on the lower foremast platform and $\text{SST}_{5\text{m}}$ were measured with a SeaBird SBE21 thermosalinograph installed in the sea surface water monitoring laboratory in the bow of the ship. Wind speeds were measured with a Koshin Denki KE-500 anemometer mounted at the top of the foremast at a height of 30 m.

d. *Results*

For each dataset, the mean ΔT_{depth} and wind speed has been computed at 10-min intervals along the tran-

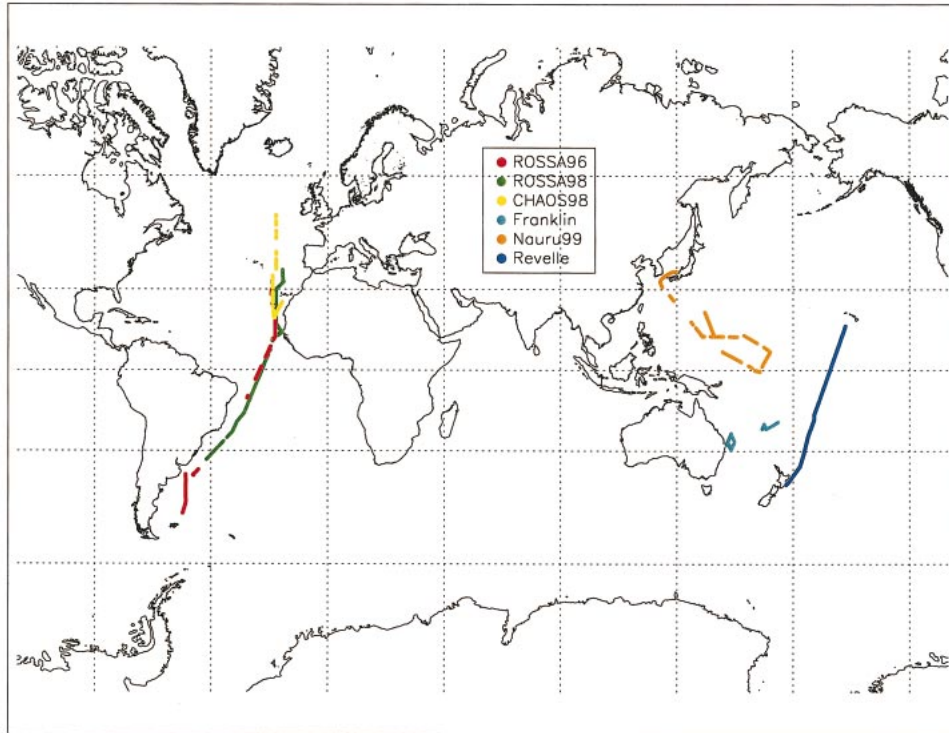


FIG. 3. Location of the in situ observations used in this study.

sects shown in Fig. 3 to match the sampling interval of the M-AERI system. The ΔT_{depth} for each transect was then plotted as a function of measured wind speed and presented in Figs. 4a–f. Each average data point has been color coded according to the local time of data acquisition in order to distinguish between nighttime and daytime data. Inspection of Fig. 4 reveals small slopes in ΔT_{depth} for the “higher” wind speed regime ($>6 \text{ m s}^{-1}$) having mean values of $-0.011 \text{ K (m s}^{-1})^{-1}$ for nighttime data and $+0.0045 \text{ K (m s}^{-1})^{-1}$ for daytime data (from Table 1). This could be due to a number of factors but is probably due to volume warming of the skin layer and subskin layer relative to the deeper $\text{SST}_{\text{depth}}$ measurement. Figure 4 also shows a small number of outlying data points in the higher wind speed domain. The accuracy of a radiometric SST_{skin} measurement depends on a number of issues including an adequate correction for the downwelling radiance emanating from the atmosphere, reflected at the sea surface (due to the nonunity emissivity value of seawater) into the in situ radiometer field of view and surface foam and waves. As shown by Donlon and Nightingale (2000), limitations in the traditional technique (near-contemporaneous sea and sky observations) used by most in situ radiometer systems to correct for sky reflections together with the effect of surface roughness limits the accuracy of a radiometric in situ SST_{skin} measurements. For partially cloudy skies the expected variability is in the order of $\pm 0.1 \text{ K}$ irrespective of the

radiometer accuracy and is clearly seen in Fig. 4 for cases of $u > 6 \text{ m s}^{-1}$. We suggest that unexpected warm ΔT_{depth} data are probably due to an overcompensation of sky radiance reflection while cool ΔT_{depth} data (in the same wind speed regime) are assumed to be caused by an undercompensation of sky radiance. In both cases the error is related principally to surface roughness effects (waves and swell waves reflecting different parts of the sky radiance) and ship roll and pitch movements modifying the radiometer view geometry. These effects result in sky radiance measurements that do not correspond to the actual sky radiance that is reflected at the sea surface and in a “worst case” scenario, a clear sky instead of a cloudy sky radiance is used to calculate the SST_{skin} (e.g., Katsaros and Businger 1973; Donlon and Nightingale 2000).

In the case of the M-AERI (*Revelle 97*) data, this source of uncertainty is less likely, because of the spectral interval chosen for the SST_{skin} measurements. As a result of the broad spectral range of the M-AERI data ($\sim 3\text{--}18 \mu\text{m}$), the SST_{skin} can be determined in a large selection of spectral intervals, the only constraints being imposed by the need that the atmospheric transmission be sufficiently high for the measurement made while directed at the sea surface and be dominated by oceanic emission. The obvious choice, the $10\text{--}12\text{-}\mu\text{m}$ interval, would provide a direct comparison with the measurements made in this atmospheric window by satellite radiometers (e.g., AVHRR and ATSR). However, this

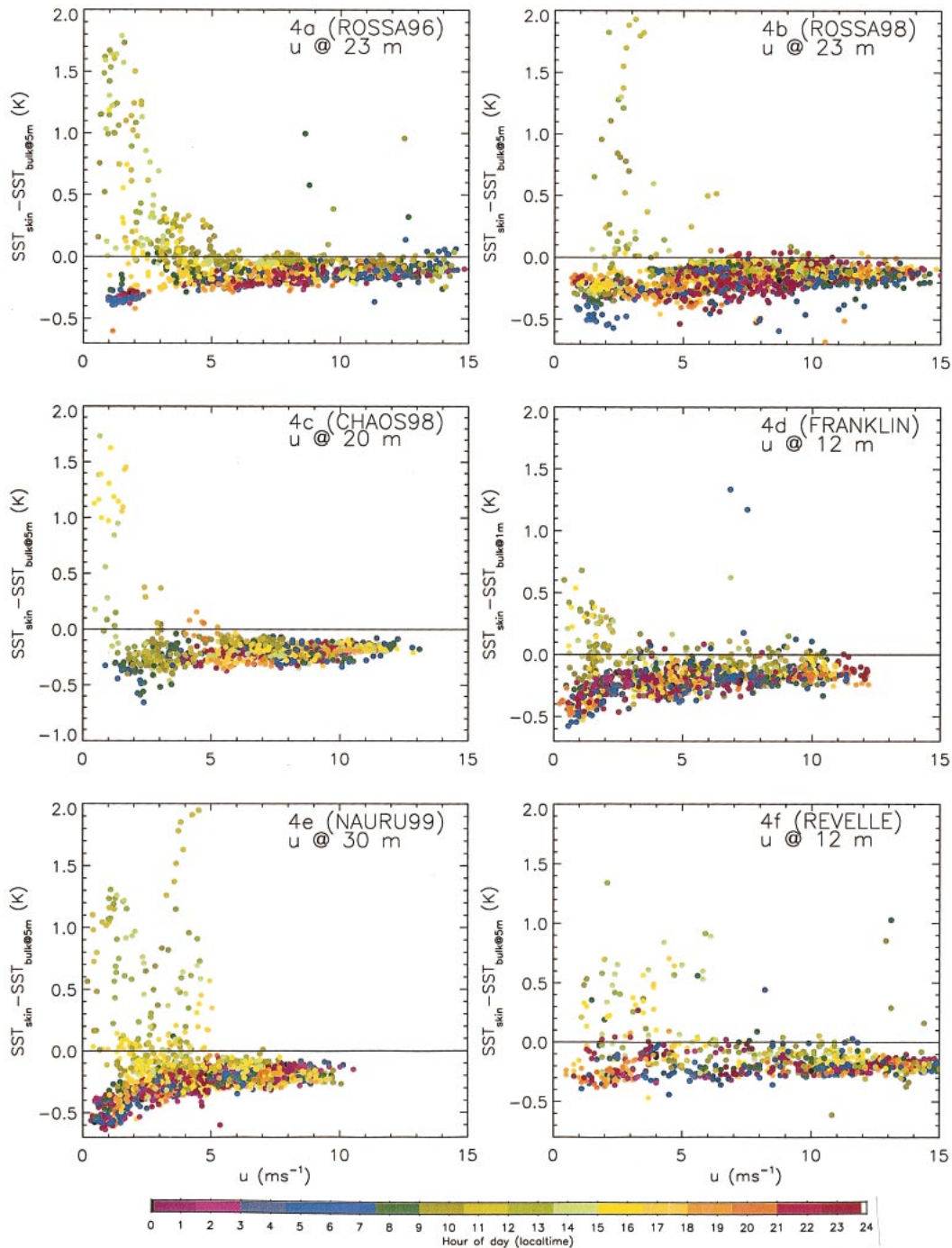


FIG. 4. The ΔT_{depth} as a function of wind speed. Each point is a 10-min mean measurement and has been colored according to the local time of day at the mean acquisition time (a) ROSSA 1996 $\Delta T_{\text{depth } 5 \text{ m}}$ as a function of wind speed at 23 m. (b) ROSSA 1998 $\Delta T_{\text{depth } 5 \text{ m}}$ as a function of wind speed at 23 m. (c) CHAOS 1998 $\Delta T_{\text{depth } 5 \text{ m}}$ as a function of wind speed at 20 m. (d) R/V *Franklin* $\Delta T_{\text{depth } 1 \text{ m}}$ as a function of wind speed at 12 m. (e) Nauru 99 $\Delta T_{\text{depth } 5 \text{ m}}$ as a function of wind speed at 30 m. (f) R/V *John-Revelle* $\Delta T_{\text{depth } 5 \text{ m}}$ as a function of wind speed at 12 m.

interval is chosen for measurements at satellite height because of the high atmospheric transmission, and may not be the best criterion for selecting the spectral interval for the short pathlengths (~ 10 m) used by the M-AERI

on the deck of a research ship. The disadvantage of using the high-transmissivity part of the spectrum is that the correction for the component of the measurement due to sky radiance reflected at the sea surface is im-

TABLE 1. Summary of $\Delta T_{\text{depth } 5 \text{ m}}$ ($^*SST_{\text{depth } 1 \text{ m}}$) characteristics for wind speed greater than 6 m s^{-1} for each dataset shown in Fig. 4.

	$\Delta T_{\text{depth } \text{day}}$				$\Delta T_{\text{depth } \text{night}}$			
	No.	Mean (K)	Std dev (K)	Slope $[\text{K} (\text{m s}^{-1})^{-1}]$	No.	Mean (K)	Std dev (K)	Slope $[\text{K} (\text{m s}^{-1})^{-1}]$
ROSSA 1996	386	-0.07	0.08	-0.004	339	-0.14	0.05	0.01
ROSSA 1998	677	-0.13	0.06	-0.002	863	-0.15	0.08	0.009
CHAOS 1998	535	-0.19	0.05	0.001	327	-0.19	0.04	0.011
R/V <i>Franklin</i> 1999*	204	-0.11	0.15	-0.014	205	-0.18	0.10	0.017
Nauru 99	747	-0.18	0.04	-0.006	609	-0.20	0.03	0.018
R/V <i>Revelle</i> 1997	260	-0.20	0.05	-0.002	264	-0.20	0.05	0.002
Summary total no. or mean	2909	-0.17	0.07	-0.0045	2607	-0.17	0.06	0.011

portant. By selecting measurements in a part of the spectrum where the atmosphere is less transmissive, the consequences of uncertainties in the correction for reflected sky radiance are much smaller, as the sky radiance originates low in the troposphere at temperatures close to that of the sea surface. For this reason the M-AERI SST_{skin} values used here are determined from measurements made at a narrow wavelength interval at $7.7 \mu\text{m}$ (Smith et al. 1996; Minnett et al. 2001). While the M-AERI data shown in Fig. 4f have a few outliers at high wind speed similar to those of other instruments; it is not thought that these are caused by errors in the corrections for sky radiance.

Table 1 summarizes the observational data presented in Figs. 4a–f. The consistency of results is testimony to the careful design, calibration, and deployment of the in situ radiometer instruments and, in all cases, mean differences between datasets approximate the accuracy limit of the radiometers themselves ($\sim 0.05 \text{ K}$). Table 1 suggests that the mean correction required to translate a well-calibrated SST_{depth} measurement to represent SST_{skin} is $0.17 \pm 0.07 \text{ K}$ rms. The agreement between these observations (collected using different instruments, in different oceans and seasons) considerably strengthens the conclusions of Donlon et al. (1999a): For the purpose of satellite SST validation exercises, at wind speeds greater than 6 m s^{-1} , well-calibrated subsurface SST_{depth} measurements obtained from ships and buoys, when corrected for a small cool bias, may be considered to represent the SST_{skin} . Nevertheless, we draw attention to the fact that at wind speeds less than 6 m s^{-1} , and in the absence of adequate models accounting for stratification and cool skin effects, in situ radiometric SST_{skin} measurements made by high accuracy in situ instrumentation are clearly mandatory for the validation of satellite SST data due to the significant decoupling between the SST_{skin} and SST_{depth} that is possible.

We further investigated the entire nighttime dataset in order to avoid the complicating effect of diurnal stratification. Figure 5 plots the nighttime only ΔT_{depth} as a function of wind speed for each of the transects shown in Fig. 3. The relationship between nighttime ΔT_{depth} and wind speed is well defined at wind speeds greater than 2 m s^{-1} and, using an optimal least-squares fit we found the following function to adequately represent the data:

$$\Delta T_{\text{depth}} = -0.14 - 0.30 \exp\left(-\frac{u}{3.7}\right), \quad (2)$$

where ΔT_{depth} (K) is the skin temperature deviation referenced to a depth measurement of approximately 5 m and u is the wind speed (m s^{-1}). It is interesting to note that the scale wind speed of 3.7 m s^{-1} in (2) marks the onset of ubiquitous small waves on the sea surface and this may suggest the lower wind speed limit above which wind-induced shear-driven turbulent heat transfer

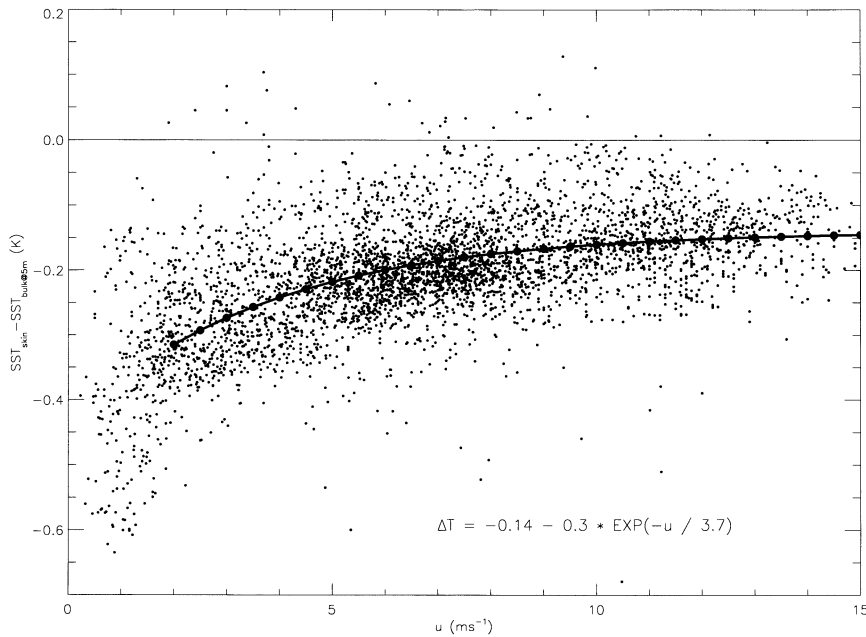


FIG. 5. All nighttime only $\Delta T_{\text{depth } 5 \text{ m}}$ data (as shown in Fig. 4) plotted as a function of wind speed.

is the governing air–sea heat transfer process. At lower wind speeds, the role of turbulent shear-driven heat transfer diminishes and convective and molecular heat transfer processes dominate when $u < 2 \text{ m s}^{-1}$ maintaining a significant temperature gradient in the skin layer (as clearly seen in Fig. 5). Consequently, the application of (2) should be treated with caution at very low ($< 2 \text{ m s}^{-1}$) wind speeds where detailed parameterizations may be more appropriate (e.g., Saunders 1967; Hasse 1971; Grassl 1976; Katsaros 1977; Fairall et al. 1996). In summary, we identify two types of in situ measurement that can be used to validate satellite-derived SST_{skin} products:

- *direct* validation measurements are those based on contemporaneous in situ SST_{skin} observations obtained using a radiometer of similar spectral characteristics to a satellite instrument and,
- *indirect* validation measurements based on well-calibrated contemporaneous in situ $\text{SST}_{\text{depth}}$ observations at a specified depth when wind speeds are greater than 6 m s^{-1} .

4. Requirements for long-term global satellite SST_{skin} validation

The purpose of comparing in situ SST_{skin} measurements with satellite-derived SST_{skin} is to validate the atmospheric correction technique and algorithm. Thus, the number and quality of comparisons are determined by the properties and the effects of the atmosphere, which is assumed to be clear of clouds. The accuracy

requirement can be set at about 0.1 K or better, given the magnitude of ΔT_{depth} and the typical accuracy requirements for long-term climate research (World Climate Research Programme 1985). The number of contemporaneous measurements is harder to establish, but one approach is to consider the variance in brightness temperature within a given “climate region” in a given time interval. Minnett (1986) found that approximately 100 randomly selected atmospheric profiles (radiosondes) were required to produce a stable estimate of rms retrieval error for a typical $\text{pSST}_{\text{depth}}$ algorithm over the North Atlantic area. A significantly smaller number would not sample the full range of atmospheric variability potentially introducing errors in the atmospheric correction process.

a. Direct validation measurements

The best satellite SST_{skin} validation measurements (although not necessary the most practical) are contemporaneous satellite and in situ SST_{skin} over a wide geographical area encompassing a range of ocean–atmosphere. While the widespread deployment of accurate ship-mounted research instrumentation such as SISTeR, M-AERI, or the DAR011 from research vessels will continue to provide a limited source of high accuracy in situ SST_{skin} measurements, the number of successful satellite SST_{skin} validation instances has so far, been relatively small. This is due to the availability and difficulty of developing, maintaining, and deploying such instrumentation on suitable research ships that need to

be positioned in clear sky conditions within the satellite swath.

There is also a logistical and financial limit to the number of in situ infrared radiometer systems that can be deployed to provide a direct SST_{skin} validation network. Following the previous discussion, given approximately 60 possible overpasses of a wide-swath satellite sensor (such as MODIS or AVHRR) in a month (assuming no swath overlap, as occurs at high latitudes, and assuming there is no need to treat day and night passes independently), we require approximately 1.5 in situ radiometer systems operating continuously in the field to populate the atmospheric parameter space for a single climatic regime. However, because only about 10% of all possible overpasses are likely to pass stringent cloud screening, this number should be increased to 15 in situ radiometer systems. Assuming three or four distinct climate regimes (e.g., high-, mid-, and low-latitude zones), the number of required radiometers becomes approximately 45, with some reduction at high latitudes for the possibility of swath overlap. If we require only seasonal, rather than monthly, validation comparisons, the number of radiometers reduces to 15. Note that if validation of narrow-swath instruments, such as ATSR on ENVISAT, is to be contemplated, this number must be increased by factor of 4 or 5. Compositing data from the same season from successive years will allow the number of radiometers to be reduced but at the cost of reliable temporal information over the lifetime of the satellite sensors.

From this simple analysis it is clear that the limited number of in situ radiometer and spectroradiometer systems both available and planned for the near future, will be unable to meet this target. This identifies a need to shift from a small number of highly sophisticated and dedicated scientific measurement campaigns to the use of reasonably accurate (~ 0.1 K rms), but less sophisticated (and therefore lower cost) autonomous instrumentation that can be widely deployed from opportunistic platforms and ships. There has been a positive development of low-cost infrared radiometer systems suitable for extended autonomous operation has focused on the use of filter radiometers (e.g., Donlon et al. 1998; Keogh et al. 1999; Jessup 2000). The need for traceability of calibration and formal protocols for the deployment of all in situ instrumentation is a prerequisite to ensure the stability and reliability of these data. Several experiments have already highlighted the considerable benefits of these exercises (Donlon et al. 1999b; Kanneburg 1998; P. J. Minnett et al. 2001, unpublished manuscript). Recalling that the use of in situ SST_{depth} measurements are not appropriate for low wind speeds, the deployment of an operational autonomous infrared radiometer measuring SST_{skin} from merchant vessels operating along regular meridional ocean-basin passages targeting low-wind speed areas is a fundamental component of a successful direct SST_{skin} validation strategy. Observations from such systems will provide long-term

accurate baseline validation data complemented by accurate measurements obtained from research instrumentation deployed during dedicated campaigns.

b. Indirect validation using SST_{depth} measurements

The fact that only a relatively small number of direct in situ SST_{skin} validation data are available has perpetuated the use of in situ SST_{depth} observations derived from widely available and globally distributed ship and buoy measurements as a primary satellite SST_{skin} validation data source. In fact this has been one of the major driving forces behind the continued use of the p SST_{depth} atmospheric correction strategy. Considering the data presented in Fig. 4, there is considerable merit in this approach providing that the in situ data are carefully screened for higher wind speed conditions only and corrected for a cool skin temperature bias. Additional benefits of this approach include 1) vertical and horizontal SST gradients will be limited within a satellite pixel due to wind mixing of the upper layers homogenizing the temperature structure, so minimizing discrepancies between point (from a buoy) or line source (from a ship) and spatially averaged measurements (from a satellite); 2) Different satellite sensors and SST data products may be compared with each other without the concern of conflicting SST definitions; 3) the considerable investment already committed to existing oceanographic infrastructure can, in principle (after rigorous quality control), be confidently used to validate satellite SST data; and 4) the use of dedicated research in situ instrumentation can be better targeted to areas characterised by low wind speeds in which the SST_{depth} approach is not expected to perform well.

Considering the large number of in situ thermometers [including those on moored buoys, drifting buoys, and volunteer observing ships (VOSs)] that are capable of measuring SST, and following the same analysis as for direct validation measurements, it would seem a simple matter to collocate wide or even narrow swath satellite sensors. However, the recent work of Kearns et al. (2000) suggests that it is still a challenge to obtain regular long-term, quality-controlled contemporaneous in situ and satellite observations. Furthermore, in many cases, the quality of indirect measurements is poor due to calibration uncertainties associated with in situ sensors, their installation and operation (Taylor et al. 1998; Emery et al. 2001). Finally, contemporaneous wind speed data are also required if these data are to form the major input to an indirect SST_{skin} validation strategy, further limiting the available dataset.

1) WIND SPEED LIMITATIONS TO THE INDIRECT SST_{skin} VALIDATION TECHNIQUE

We use data from the Special Sensor Microwave Imager (SSM/I) to define areas where indirect SST_{skin} validation should be targeted. Figure 6 shows probability

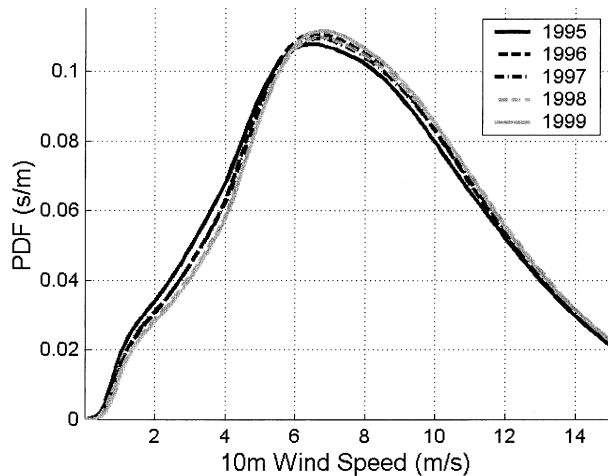


FIG. 6. Annual global F-13 SSM/I 10-m wind speed PDFs for 1995–99.

density function (PDF) computed from the annual global F-13 SSM/I 10-m wind speed for the years 1995–99. The 1995–97 PDFs show a slightly larger probability of winds less than 6 m s^{-1} than the 1998 and 1999 PDFs. The average wind speed is 8.3 m s^{-1} and approximately 30.0% of all winds are less than or equal to 6 m s^{-1} . Three percent of all winds are less than 2 m s^{-1} , which are the conditions promoting exceptionally strong diurnal stratification. Figure 7 shows the annual, global distribution of 10-m wind speeds less than 6 m s^{-1} derived from F-13 SSM/I for 1995–99. Dark regions indicate areas frequently characterized by low winds throughout the year and are expected to be characterized by large daytime ΔT values. In these areas, indirect SST_{skin} validation strategies are unlikely to provide any useful data but instead, direct radiometer measurements

should be used. Lighter areas indicate wind speeds characteristically greater than 6 m s^{-1} , indicating well-mixed surface ocean regions. Clearly large regions of the global ocean are characterized by wind speed regimes that are, in principle, appropriate for indirect SST_{skin} validation strategies. Furthermore, the global prevalence of wind speeds greater than 6 m s^{-1} is, in part, why $\text{p-SST}_{\text{depth}}$ regression type algorithms have provided useful results over the past two decades.

2) IN SITU $\text{SST}_{\text{depth}}$ OBSERVATIONS LIMITING THE INDIRECT SST_{skin} VALIDATION TECHNIQUE

The accuracy, and therefore usefulness of in situ $\text{SST}_{\text{depth}}$ data within an indirect SST_{skin} validation strategy is, in addition to the behavior of the deployment platform (buoy design, ship, etc.), critically dependent on adequate knowledge of sensor calibration stability and drift. Emery et al. (1997) showed that the depth of $\text{SST}_{\text{depth}}$ measurement can vary significantly depending on ship loading whereas Bitterman and Hanson (1993) show several designs of surface drifter that locate $\text{SST}_{\text{depth}}$ sensors in different depths. We note that the indirect SST_{skin} validation method is concerned with well-mixed conditions so that the latter issues are only important at exceptionally high wind speeds when the vertical position of $\text{SST}_{\text{depth}}$ sensors can change as larger drifters may be blown over or adversely affected by the wave field.

There are however, considerable calibration problems demonstrated by differences between $\text{SST}_{\text{depth}}$ data collected from buoys and ships using different methods and sensors (engine room intake, bucket temperature, and hull contact) that are well reviewed by Kent et al. (1991, 1993). In particular, many engine intake tem-

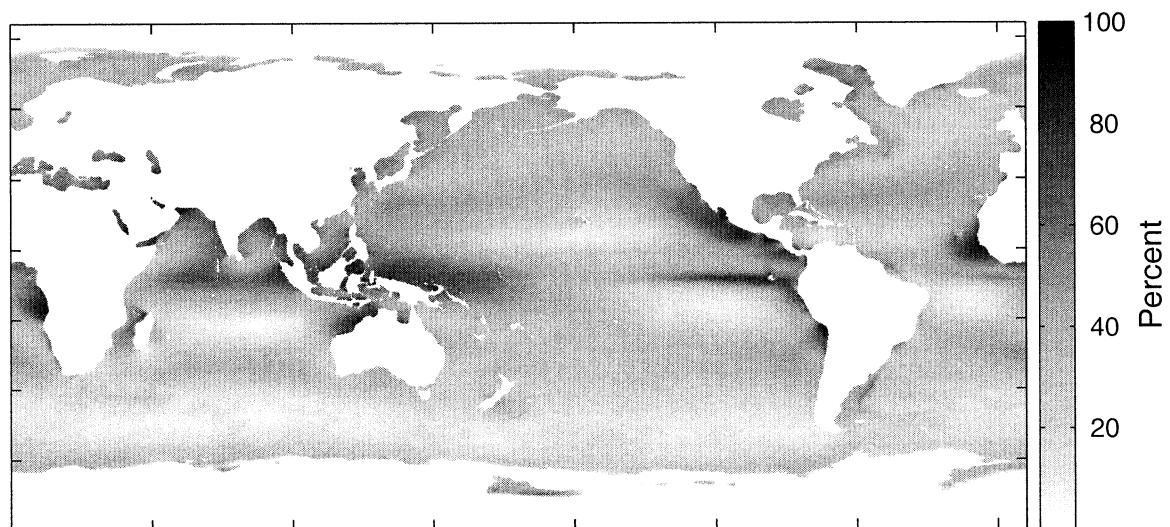


FIG. 7. Annual percentage of 10-m wind speeds less than 6 m s^{-1} from mean composite F-13 SSM/I 10-m wind speed observations for the period 1995–99. Dark regions indicate areas frequently characterized by low winds throughout the year. Lighter regions are characterized by infrequent low wind speeds.

perature data are poor in quality. In contrast, a new generation of hull-mounted sensors, if properly calibrated and maintained, are capable of providing high-quality SST_{depth} data (Taylor et al. 1998). Higher quality SST_{depth} data can be derived also from drifting and moored buoy sensors (e.g., Podesta et al. 1995; Kumar et al. 2000; Kilpatrick et al. 2001) although careful attention is required to ensure that calibration is maintained throughout the lifetime of the deployment.

SST_{depth} sensor calibration uncertainty remains the largest obstacle to the operational use of the indirect SST_{skin} validation method if a high-quality validation is required. For each particular SST_{depth} data record, a traceable end-to-end calibration should be available together with the details of the deployment platform characteristics. In this way, standards and quality control procedures can be introduced and refined improving the quality of individual systems and collectively providing a robust foundation to global datasets. Given these issues, it would be wise to consider the development of a dedicated network of SST_{depth} observations for use in a long-term indirect satellite SST_{skin} validation program.

3) INDIRECT SST_{skin} VALIDATION OF ERS-1 ALONG TRACK SCANNING RADIOMETER SST_{skin}

We applied the indirect SST_{skin} validation technique to ERS-1 ATSR measurements together with data obtained from the Tropical Ocean Global Atmosphere (TOGA) Tropical Atmosphere Ocean (TAO) project buoy array. On average, the TAO buoy arrays have winds greater than 6 m s⁻¹ for approximately 57% of the year based on F-10 SSM/I 10-m wind speed analysis. While restricted to a particular climatic region (tropical), the well-calibrated SST_{depth} observations made by these moorings allow us to test the application of indirect SST_{skin} validation.

Extending the work of Murray et al. (2000), ΔT_{depth} was computed for the first 4 years of the ATSR-1 mission from contemporaneous (± 1 h) ATSR-1 10-arc-minute (~ 18 km \times 18 km) SST_{skin} data and SST_{depth} derived from the TAO moorings at a nominal depth of 1 m. ATSR SST_{skin} data were derived using the dual-view algorithm proposed by Merchant et al. (1999), modified to account for the ATSR detector temperature variation. The mean of a 4 \times 4 array of ATSR 10-arc-minute SST_{skin} data centered over the TAO mooring was used in this analysis. All ATSR data were stringently filtered to exclude cloud-contaminated data and areas characterized by highly dynamic SST_{skin} that would otherwise complicate the analysis. This procedure resulted in approximately 7000 observations (combined day- and nighttime) that were further quality controlled to ensure the highest possible collocation in space and time. These have been plotted as a function of TAO wind speed (adjusted to 10-m height) in Fig. 8a. Also shown are the mean ΔT_{depth} values for wind speed intervals of 0.5 m s⁻¹ and error bars showing ± 1 standard deviation.

The mean daytime results highlight the complicating dominance of thermal stratification and suggest that diurnal warming occurs up to a wind speed of approximately 7 m s⁻¹. The mean difference between ATSR SST_{skin} and TAO SST_{depth} when wind speed is greater than 6 m s⁻¹ is -0.17 ± 0.46 K rms for nighttime conditions and 0.07 ± 0.47 K rms for daytime conditions and are remarkably similar to the in situ observations previously discussed. While we acknowledge the small number of data at the higher wind speed regime, after applying a cool-bias correction of 0.17 K [or using (2)] to the TAO data the mean nighttime difference is approximately zero, demonstrating the exceptional accuracy of the ATSR-1 SST_{skin} products in the tropical Pacific Ocean. This is a considerable achievement, not only because of the heavy water vapor-loading characteristic of the tropical atmosphere, but more importantly, during the analysis period, the atmosphere was heavily laden with volcanic aerosols following the eruption of Mount Pinatubo. The ATSR instrument is less affected by volcanic aerosol compared to other satellite sensors due to the along-track-scanning technique providing a direct measure of atmospheric attenuation. The data presented do demonstrate a large standard deviation of about 0.5 K that we believe is principally related to the spatial and temporal mismatch between buoy and satellite observations. Nevertheless, given the small temporal differences (± 1 h) between the point measurement of the TAO moorings and the considerably larger ATSR footprint (about 1296 km² if all four 10-arc-minute cells are included in the analysis), the agreement between the mean observations shown here and those of Fig. 4 is extremely encouraging.

Although this comparison is regional in scope, alternative SST_{depth} platforms (e.g., drifting buoys and VOSs) could be used increasing coverage and allowing a more complete validation exercise to be conducted for different climatic regions. A particular benefit of using this type of existing observational infrastructure is that remote areas characterized by high wind speed such as the Southern Ocean (where direct validation data are extremely scarce) may be included in validation studies.

4) INDIRECT SST_{skin} VALIDATION OF PATHFINDER AVHRR SST

We also applied the indirect SST_{skin} validation technique to the AVHRR Pathfinder SST (MPSST) (Evans and Podesta 1998b) in the tropical Pacific Ocean using the data provided in the University of Miami Pathfinder Match-up Database (PFMDB; Evans and Podesta 1998a). Within the PFMDB, stringent time-space match-up constraints are used to define valid contemporaneous satellite and TAO in situ measurements of within ± 1 h of satellite data acquisition and $\pm 0.1^\circ$ of latitude and longitude. All satellite observations have been quality controlled to identify and remove cloud-contaminated data from the database. The MPSST is a

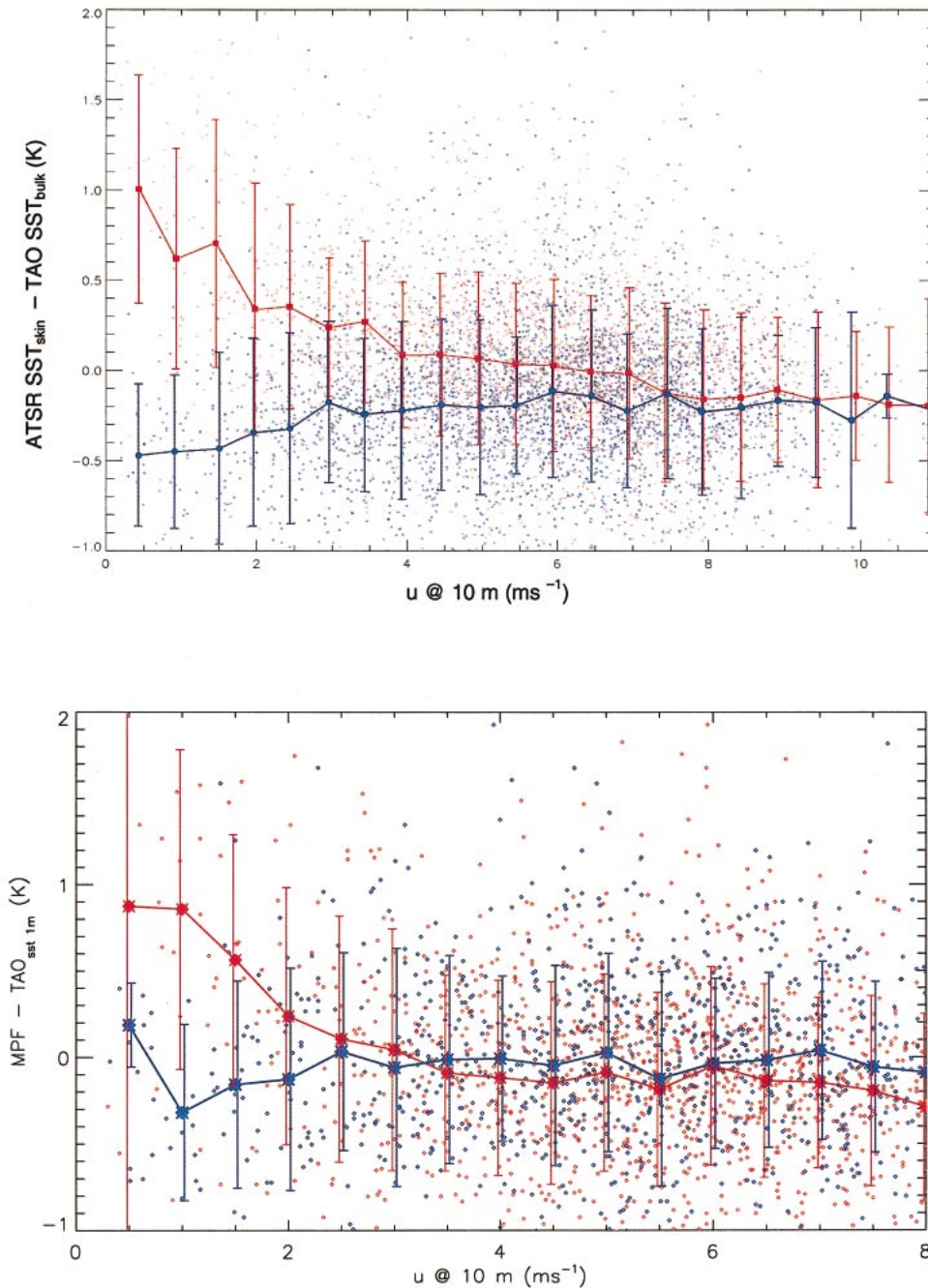


FIG. 8. (a) ATSR-1 10-arc-minute $\text{SST}_{\text{skin}} - \text{TAO SST}_{\text{depth } 1 \text{ m}}$ for the period 1992–95 plotted as a function of TAO wind speed corrected to 10-m height. Red plots consider only daytime observations and blue plots only nighttime observations. Error bars show ± 1 std dev. (b) *NOAA-14* AVHRR Miami Pathfinder $\text{SST} - \text{TAO SST}_{\text{depth } 1 \text{ m}}$ for 1998 plotted as a function of TAO wind speed corrected to 10-m height. Red plots consider only daytime observations and blue plots only nighttime observations. Error bars show ± 1 std dev.

semiempirical $\text{pSST}_{\text{depth}}$ algorithm based on separate monthly regression (thereby accounting for gross atmospheric change associated with injection and subsequent decay of volcanic aerosol, e.g.) of globally distributed contemporaneous AVHRR and in situ $\text{SST}_{\text{depth}}$ observations. Unlike the ATSR SST_{skin} data discussed

previously, the MPFSST algorithm attempts to estimate the $\text{SST}_{\text{depth}}$ and thus, comparison to in situ $\text{SST}_{\text{depth}}$ in the higher wind speed regime should result in a zero bias. The MPFSST algorithm includes $\text{SST}_{\text{depth}}$ from TAO moorings as well as those from drifting buoys, National Data Buoy Center (NDBC) and Pilot Research

Moored Array in the Tropical Atlantic (PIRATA) moored buoys in the derivation of the $pSST_{depth}$ algorithm. The regression is derived from all these data so that, mean differences found in comparisons with buoy retrievals are expected to be small. Unlike the comparison between ATSR and TAO measurements, that of MPFSST and TAO data is not between strictly independent data sets. However, the number of data points from TAO moorings used in the MPFSST algorithm derivation is a small fraction of the total (e.g., 8.6% in 1997; Kilpatrick et al. 2001), so the analysis still has merit. Furthermore, we are not drawing conclusions about rms error but about how day and night ΔT_{depth} changes with wind speed.

We extracted all TAO mooring match-ups with the National Oceanic and Atmospheric Administration *NOAA-14* AVHRR data from the PFMDB where a valid wind speed was available for 1998 and 1999. ΔT_{depth} for both day- and nighttime conditions have been computed using $\Delta T_{depth} = MPFSST - TAO$ and plotted in Fig. 8b together with the mean ΔT_{depth} for 0.5 m s^{-1} intervals and error bars describing ± 1 standard deviation. Although interesting differences between these results and those of Fig. 4 are apparent, we note the sparsity of data at lower and higher wind speeds. In particular, a divergence between the nighttime and daytime data occurs at a significantly lower wind speed of about 2.5 m s^{-1} . At wind speeds higher than this, daytime data demonstrate a cool (-0.15 K) bias with respect to the nighttime data that have a negligible ΔT_{depth} of $-0.06 \text{ K} \pm 0.5 \text{ K}$ rms. At wind speeds lower than 2.5 m s^{-1} , a ΔT_{depth} of similar magnitude to the in situ data shown in Fig. 4 is apparent in the nighttime data and the effect of diurnal stratification is apparent in the daytime data. These data confirm the ability of the MPFSST algorithm to estimate the SST_{depth} at wind speeds greater than 2.5 m s^{-1} at nighttime, although when wind speed $< \sim 2.5 \text{ m s}^{-1}$, a cool skin effect persists.

5. Discussion and conclusions

The data presented here highlight the need to introduce and maintain a rigorous approach to the validation of satellite-derived SST_{skin} , based on a combination of existing oceanographic infrastructure (buoys and ships), research-grade radiometers (e.g., DAR011, SISTeR, MAERI) and emerging autonomous instrumentation. This is required for the immediate validation of satellite SST products but also, for the future reanalysis and synergistic development of multisensor data products. It is fundamental to the development of long-term multisensor SST time series that accurate and continuing validation of the satellite sensors and associated data products is accomplished in order to demonstrate confidence in the data merging procedures and cross calibration of the sensors themselves (Donlon et al. 1999). An inability to accurately and repeatedly quantify appropriate confidence limits for satellite data using the current oper-

ational in situ infrastructure alone (e.g., buoy installations) is very apparent (e.g., Donlon and Robinson 1997). We have shown that the use of subsurface SST_{depth} observations can provide a satisfactory SST_{skin} validation data source when wind speed conditions are greater than 6 m s^{-1} and a small bias correction is applied to the SST_{depth} data. However, considerable effort is still required to ensure and maintain adequate sensor calibrations and to provide adequate data coverage in the most appropriate areas of the ocean and poses the greatest limitation on the application of indirect SST_{skin} validation. In particular, we believe that these issues should be incorporated into a new generation of dedicated SST_{depth} measurement infrastructure targeted at the continuous long-term validation of satellite-derived SST_{skin} data. We suggest that the use of recently developed autonomous, instrumentation that can be widely deployed on ships of opportunity operating along regular routes could provide a new approach. However, these instruments should be considered in a more holistic framework: there is no substitute for the exceptionally widespread (both in time and space) high-quality SST_{depth} measurements provided by buoys and ships if used carefully and appropriately to validate satellite SST observations.

As the need for a new generation of satellite SST product is emerging based on merged multisensor SST data, it is important that the benefits of the indirect SST_{skin} validation method are fully realized. For example, infrared and microwave SST estimates represent the SST_{skin} and $SST_{subskin}$, respectively; only the indirect SST_{skin} validation method will provide a robust means to directly compare observations made by several satellite instruments to investigate systematic differences. The indirect validation method provides a means to retrospectively generate and validate historical satellite archives for which no direct SST_{skin} observations are available. Furthermore, the method provides a simple way to reconcile differences between SST_{skin} and SST_{depth} , facilitating the direct use of satellite SST data within Numerical Weather Prediction models that hitherto, have been degraded through techniques that constrain satellite SST observations to in situ observations (e.g., Kara et al. 2000). Without a coordinated effort to increase the quality, access, and availability of in situ measurements representative of the global atmosphere and ocean, the considerable financial investment committed to earth observation will be compromised by poor quality cross- and intercalibration of satellite and in situ instruments and/or insufficient validation of data products. We suggest that SST_{depth} calibration records and documentation are coordinated through a dedicated program focusing on the validation of satellite-derived SST. Such a framework can only be achieved through international collaboration with several research groups and national agencies.

Acknowledgments. To the captains and companies of the RRS *James Clark Ross*, R/V *Franklin*, R/V *Revelle*, RRS *Discovery*, and the R/V *Mirai*. PJM acknowledges support from NASA through Grant NAG56577. Special thanks to Tom Sheasby for permission to use the CHAOS data. The ATSR data are provided by courtesy of the European Space Agency and have been analyzed by the Rutherford Appleton Laboratory, Chilton, Didcot, Oxon, United Kingdom.

REFERENCES

- Barton, I. J., 1995: Satellite derived sea surface temperatures: Current status. *J. Geophys. Res.*, **100**, 8777–8790.
- , 1998: Future directions for satellite studies of skin-depth temperature differences. *J. Adv. Mar. Sci. Tech. Soc.*, **4**, 197–204.
- , 2001: Interpretation of satellite-derived sea surface temperatures. *Adv. Space Res.*, **28**, 165–170.
- Bitterman, D. S., and D. V. Hansen, 1993: Evaluation of sea surface temperature measurements from drifting buoys. *J. Atmos. Oceanic Technol.*, **10**, 88–96.
- Donlon, C. J., 2001: The global ocean data assimilation experiment high resolution sea surface temperature pilot project strategy and implementation plan. Marine Environment Unit, Joint Research Centre, Ispra, Italy, 58 pp.
- , and I. S. Robinson, 1997: Observations of the oceanic thermal skin in the Atlantic Ocean. *J. Geophys. Res.*, **102**, 18 585–18 606.
- , and T. J. Nightingale, 2000: The effect of atmospheric radiance errors in radiometric sea surface temperature measurements. *Appl. Opt.*, **39**, 2387–2392.
- , and Coauthors, 1998: Solid state radiometer measurements of sea surface skin temperature. *J. Atmos. Oceanic Technol.*, **15**, 774–776.
- , T. J. Nightingale, T. Sheasby, J. Turner, I. S. Robinson, and W. J. Emery, 1999a: Implications of the oceanic thermal skin temperature deviation at high wind speeds. *Geophys. Res. Lett.*, **26**, 2505–2508.
- , —, L. Fiedler, G. Fisher, D. Baldwin, and I. S. Robinson, 1999b: A low cost blackbody for the calibration of sea going infrared radiometer systems. *J. Atmos. Oceanic Technol.*, **16**, 1183–1197.
- Emery, W. J., K. Cherkauer, B. Shannon, and R. W. Reynolds, 1997: Hull-mounted sea surface temperatures from ships of opportunity. *J. Atmos. Oceanic Technol.*, **14**, 1237–1251.
- , D. J. Baldwin, P. Schlüssel, and R. W. Reynolds, 2001: Accuracy of in situ sea surface temperatures used to calibrate infrared satellite measurements. *J. Geophys. Res.*, **106**, 2387–2405.
- Esaias, W. E., and Coauthors, 1998: An overview of MODIS capabilities for ocean science observations. *IEEE Trans. Geosci. Remote Sens.*, **36**, 1250–1265.
- Evans, R., and G. Podesta, 1998a: *Matchup Database 1985–1997 (Version 19.0)*. [Available from the University of Miami, RSMAS/MPO, 4600 Rickenbacker Causeway, Miami, FL 33149-1098.]
- , and —, 1998b: *Pathfinder Sea Surface Temperature Algorithm Version 4.0*. [Available from the University of Miami, RSMAS/MPO, 4600 Rickenbacker Causeway, Miami, FL 33149-1098.]
- Fairall, C., E. Bradley, J. Godfrey, G. Wick, J. Edson, and G. Young, 1996: Cool-skin and warm-layer effects on sea surface temperature. *J. Geophys. Res.*, **101**, 1295–1308.
- Grassl, H., 1976: On the dependence of the measured cool skin of the ocean on wind stress and total heat flux. *Bound.-Layer Meteor.*, **10**, 465–474.
- Hasse, L., 1971: The sea surface temperature deviation and the heat flow at the air–sea interface. *Bound.-Layer Meteor.*, **1**, 368–379.
- Jessup, A. T., 2000: CIRIMS: Calibrated infrared in situ measurement system. *Proc. IEEE Int. Geosciences and Remote Sensing Symp.*, Hamburg, Germany, IGARSS.
- Kannenburg, R., 1998: IR instrument comparison workshop at RSMAS. *Earth Observer*, **10**. [Available online at http://eosps.gsf.nasa.gov/eos_observ/5_6_98/p51.html.]
- Kara, A. B., P. A. Rochford, and H. E. Hurlburt, 2000: Efficient and accurate depth parameterizations of air–sea fluxes for use in general circulation models. *J. Atmos. Oceanic Technol.*, **17**, 1421–1438.
- Katsaros, K. B., 1977: The surface temperature deviation at low wind speeds: Is there a limit? *Tellus*, **29**, 229–239.
- , and J. A. Businger, 1973: Comments on the determination of the total heat flux from the sea with a two wavelength radiometer system developed by McAlister. *J. Geophys. Res.*, **78**, 1964–1970.
- Kearns, E. J., J. A. Hanafin, R. H. Evans, P. J. Minnett, and O. B. Brown, 2000: An independent assessment of Pathfinder AVHRR sea surface temperature accuracy using the Marine Atmosphere Emitted Radiance Interferometer (MAERI). *Bull. Amer. Meteor. Soc.*, **81**, 1525–1536.
- Kent, E. C., B. S. Truscott, J. S. Hopkins, and P. K. Taylor, 1991: The accuracy of ship’s meteorological observation—Results of the VSOP-NA. Marine Meteorology and Related Oceanographic Activities Rep. 26, World Meteorological Organization, Geneva, Switzerland, 86 pp.
- , P. K. Taylor, B. S. Truscott, and J. A. Hopkins, 1993: The accuracy of voluntary observing ship’s meteorological observations. *J. Atmos. Oceanic Technol.*, **10**, 591–608.
- , T. Forrester, and P. K. Taylor, 1996: A comparison of the oceanic skin effect parameterisations using ship borne radiometer data. *J. Geophys. Res.*, **101**, 16 649–16 666.
- Keogh, S. J., I. S. Robinson, C. J. Donlon, and T. J. Nightingale, 1999: The validation of AVHRR SST using shipborne radiometers. *Int. J. Remote Sens.*, **20**, 2871–2876.
- Kilpatrick, K. A., G. P. Podesta, and R. H. Evans, 2001: Overview of the NOAA/NASA Pathfinder algorithm for sea surface temperature and associated matchup database. *J. Geophys. Res.*, **106**, 9179–9198.
- Kumar, A., P. J. Minnett, G. Podesta, and R. Evans, 2000: An analysis of Pathfinder SST algorithm for global and regional conditions. *Proc. Indian Acad. Sci.: Earth Planet. Sci.*, **109**, 395–405.
- Kummerow, C., W. Barnes, T. Kozu, J. Shiue, and J. Simpson, 1998: The Tropical Rainfall Measuring Mission (TRMM) sensor package. *J. Atmos. Oceanic Technol.*, **15**, 809–817.
- Le Traon, P.-Y., M. Rienecker, N. Smith, P. Baharel, M. Bell, H. Hurlburt, and P. Dandin, 1999: Operational oceanography and prediction, a GOADE perspective. *Ocean Obs Conf.*, San Raphael, France, 21.
- Llewellyn-Jones, D. T., P. J. Minnett, R. W. Saunders, and A. M. Zavody, 1984: Satellite multichannel infrared measurements of sea-surface temperature of the N.E. Atlantic Ocean using AVHRR/2. *Quart. J. Roy. Meteor. Soc.*, **110**, 613–631.
- McKeown, W., and W. Asher, 1997: A radiometric method to measure the concentration boundary layer thickness at an air–water interface. *J. Atmos. Oceanic Technol.*, **14**, 1494–1501.
- , F. Bretherton, H. L. Huang, W. L. Smith, and H. L. Revercomb, 1995: Sounding the skin of water: Sensing air–water interface temperature gradients with interferometry. *J. Atmos. Oceanic Technol.*, **12**, 1313–1327.
- Merchant, C. J., A. R. Harris, M. J. Murray, and A. M. Zavody, 1999: Toward the elimination of bias in satellite retrievals of sea surface temperature 1. Theory, modelling and inter-algorithm comparison. *J. Geophys. Res.*, **104**, 23 565–23 578.
- Minnett, P. J., 1986: A numerical study of the effects of anomalous North Atlantic atmospheric conditions on the infrared measurement of sea-surface temperature from space. *J. Geophys. Res.*, **91**, 8509–8521.
- , 1990: The regional optimization of infrared measurements of sea-surface temperature from space. *J. Geophys. Res.*, **95**, 13 497–13 510.

- , and J. A. Hanafin, 1998: At-sea measurements of the ocean skin temperature and its response to surface fluxes. *IEEE Int. Geosci. Remote Sens. Symp.*, Seattle WA, IEEE.
- , and B. Ward, 2000: Measurements of near-surface ocean temperature variability—Consequences on the validation of AATSR on Envisat. *Proc. ERS-ENVISAT Symp. "Looking down to Earth in the New Millennium,"* Gothenburg, Sweden, European Space Agency, 10 pp. [Available as CD-ROM from ESA Publications Division, ESTEC, P.O. Box 299, 2200 AG Noordwijk, Netherlands.]
- , R. O. Knuteson, F. A. Best, B. J. Osborne, J. A. Hanafin, and O. B. Brown, 2001: The Marine-Atmosphere Emitted Radiance Interferometer: A high-accuracy, seagoing infrared spectroradiometer. *J. Atmos. Oceanic Technol.*, **18**, 994–1013.
- Murray, M. J., M. R. Allen, C. J. Merchant, A. R. Harris, and C. J. Donlon, 2000: Direct observations of skin-depth SST variability. *Geophys. Res. Lett.*, **27**, 1171–1174.
- Mutlow, C. T., A. M. Zavody, I. J. Barton, and D. T. Llewellyn-Jones, 1994: Sea surface temperature measurements by the along-track scanning radiometer on the ERS-1 satellite: Early results. *J. Geophys. Res.*, **99**, 22 575–22 588.
- Parkes, I. M., and Coauthors, 2000: The Mutsu bay experiment—Validation of ATSR-1 and ATSR-2 sea surface temperature (SST) in Mutsu Bay, Japan during July and August 1996. *Int. J. Remote Sens.*, **21**, 3445–3460.
- Paulson, C. A., and J. J. Simpson, 1981: The temperature difference across the cool skin of the ocean. *J. Geophys. Res.*, **86**, 11 044–11 054.
- Podesta, G. P., S. Shenoi, J. W. Brown, and R. H. Evans, 1995: AVHRR Pathfinder Oceans Matchup Database 1985–1993 (Version 18). Draft tech. report of the University of Miami Rosenstiel School of Marine and Atmospheric Science, 33 pp.
- Saunders, P. M., 1967: The temperature at the ocean–air interface. *J. Atmos. Sci.*, **24**, 269–273.
- Saunders, R. W., and D. P. Edwards, 1989: Atmospheric transmittances for the AVHRR channels. *Appl. Opt.*, **28**, 4154–4160.
- Schlüssel, P., W. J. Emery, H. Grassl, and T. Mammen, 1990: On the depth skin temperature difference and its impact on satellite remote sensing of sea surface temperature. *J. Geophys. Res.*, **95**, 13 341–13 356.
- Smith, N., 2000: *GODAE High Resolution Sea Surface Temperature Workshop Proceedings*. Ispra, Italy, European Commission Joint Research Centre, Space Applications Institute, Marine Environment Unit, 90 pp.
- Smith, W. L., and Coauthors, 1996: Observations of the infrared radiative properties of the ocean—Implications for the measurement of sea surface temperature via satellite remote sensing. *Bull. Amer. Meteor. Soc.*, **77**, 41–51.
- Taylor, P. K., E. C. Kent, and S. A. Josey, 1998: The accuracy of sea surface temperature data from Voluntary Observing Ships. Report of the OOPC/AOPC Workshop on Global Sea Surface Temperature Data Sets, Palisades, NY, 2–4 November 1998, WMO, Geneva, Annex III, 51–54.
- Vasquez, J., K. Perry, and K. Kilpatrick, 1998: NOAA/NASA AVHRR Oceans Pathfinder Sea Surface Temperature Data Set; User's Reference Manual Version 4.0. JPL Publication D-14070, 5 pp. [Available online at http://podaac.jpl.nasa.gov/pub/sea_surface_temperature/avhrr/pathfinder/doc/usr_gde4.0.toc.html.]
- Ward, B., and P. J. Minnett, 2001: An autonomous profiler for near surface temperature measurements. *Gas Transfer at Water Surfaces, Geophys. Monogr.*, No. 127, Amer. Geophys. Union, 167–172.
- Wentz, F. J., and T. Meissner, 1999: Algorithm Theoretical Basis Document (ATBD) version 2 AMSR Ocean Algorithm. Remote Sensing Systems, 65 pp. [Available from Remote Sensing Systems, 438 First St., Suite 200, Santa Rosa, CA 95401. (<http://www.ssmi.com/papers/OCEANATBD.PDF>).]
- , C. L. Gentemann, D. K. Smith, and D. B. Chelton, 2000: Satellite measurements of sea-surface temperature through clouds. *Science*, **288**, 847–850.
- World Climate Research Programme, 1985: Scientific plan for the Tropical Global Ocean Programme. WCRP Publication 3, World Meteorological Organization, Geneva, Switzerland, 107 pp.
- Zavody, A. M., C. T. Mutlow, and D. T. Llewellyn-Jones, 1995: A radiative transfer model for SST retrieval for the ATSR. *J. Geophys. Res.*, **100**, 937–952.