Estimating Sea Surface Temperature From Infrared Satellite and In Situ Temperature Data

W.J. Emery, Sandra Castro
CCAR Box 431

U Colorado
Boulder, Co., 80309

G.A Wick
NOAA/ETL

325 Broadway
Boulder, Co., 80303

Peter Schluessel
EUMETSAT

Am Kavalleriesand
64295 Darmstadt,
Germany

Craig Donlon
CEC – JRC ISPRA,

Marine Environment
I-21020 Ispra
ITALY

Abstract

Sea surface temperature (SST) is a critical quantity in the study of both the ocean and the atmosphere as it is directly related to and often dictates the exchanges of heat, momentum and gases between the ocean and the atmosphere. As the most widely observed variable in oceanography, SST is used in many different studies of the ocean and its coupling with the atmosphere. We examine the history of this measurement and how this history led to today's practice of computing SST by regressing satellite infrared measurements against in situ SST observations made by drifting/moored buoys and ships. The fundamental differences between satellite and in situ SST are discussed and recommendations are made for how both data streams should be handled. A comprehensive in situ validation/calibration plan is proposed for the satellite SSTs and consequences of the suggested measurements are discussed with respect to the role of SST as an integral part of the fluxes between the ocean and the atmosphere.

Introduction

As one of easiest ocean variables to observe, sea surface temperature (SST) has a very long history in the study of the ocean and its interchange with the atmosphere. One of the most common operational applications of SST observations is as a boundary layer input and an assimilation data set for atmospheric circulation/forecast models. The correct treatment of accurate (~0.3 K) SST in these numerical models is needed to be able to understand and predict climate change. In addition, SST maps are vitally important for the commercial and sport fishing communities as well as critical input information for the fluxes of gases between the ocean and the atmosphere. As a result there is a current demand for the delivery of timely and accurate global SST maps. The only measurement system that provides this time/space coverage requirement is infrared satellite sensors.

In the past and at present infrared satellite data have been and are used to estimate the same type of SST measured first by ships and then later by drifting/moored buoys. Referred to generally as bulk SST, measurements from buoys and ships are collected at depths from 0.5 to 5 m below the sea surface. This is in contrast with infrared SST measurements which due to the high emissivity of sea water are representative only of depths of approximately 10 microns within the oceanic skin layer. Proper validation/calibration of these infrared skin SST measurements requires simultaneous "in situ" skin SST measurements made below the intervening atmosphere. At present the relatively few of these types of measurements that exist have all been collected as part of specific research cruises. This lack of in situ skin SST measurements had led to the common SST estimation practice to adjust the satellite SSTs to match a selection of buoy SSTs. This forces the satellite skin SSTs to estimate buoy bulk SSTs and ignores the physics that connect the skin and bulk SSTs. This relationship between the skin and bulk SSTs depends on the wind stress and net air-sea heat fluxes (Wick et al., 1996) which are parts of the climate system.

A key to being able to routinely compute the skin SST from infrared AVHRR imagery is the existence of a comprehensive validation program for in situ skin SST measurements. The primary purpose of this paper is to clearly state the need for such measurements that are accurate, reliable, widely spread geographically and continuous over time. We outline the challenges of such a validation program and introduce a possible solution based on instruments deployed on ships-of-opportunity and moored buoys. We will review recent developments in the study of skin SST and its relationship to the subsurface bulk SST and assess our ability to predict one from the other. In addition, we suggest using simultaneous measurements of bulk and skin SST along with coincident wind speed measurements for the estimation of the net air-sea heat flux. We believe that by adopting the recommended procedures it will be possible to improve the accuracy of the infrared SST retrievals and to better link satellite and in situ SST measurements. We will also expand and extend our knowledge of the relationship between skin and bulk SSTs and their relationships to net air-sea fluxes, thus improving our knowledge not only of heat and momentum fluxes but also of the gas fluxes such as carbon dioxide. Better knowledge of the role of the skin in these air-sea exchanges will contribute significantly to improving coupled atmosphere/ocean climate models.

History of SST Measurement

The earliest measurements of SST were from sailing vessels where the common practice was to collect a bucket of water while the ship was underway and then measure the temperature of this bucket of water with a mercury in glass thermometer. This then was a sample from the few upper tens of centimeters of the water. Modern powered ships made this bucket collection impractical and it became common practice to measure SST as the temperature of the seawater entering to cool the ship's engines. The depth of the inlet pipe varies with ship from about one meter to five meters. Called "injection temperature" (a thermistor is "injected" into the pipe carrying cooling water) this measurement is an analog reading of a round gauge recorded by hand and radioed in as part of the regular weather observations from merchant ships. Located in the warm engine room this SST measurement has been shown to have a warm bias (Saur, 1963) and is generally much nosier than buoy measurements of bulk SST (Emery et al., 2000). Some research vessels still use "bucket samples" for bulk SST measurements where the buckets are slim profile pieces of pipe with a thermometer built into it.

A modern alternative to the ship injection SSTs is to mount thermistors on the metal hull of the ship (Emery et al., 1997). These hulls are excellent conductors of heat and hence should represent the temperature just to the outside. To measure the bulk SST we want to sample near the surface but not above it. Mounted on merchant ships that load and unload cargo we found that we had to install 3-5 sensors arrayed vertically to be able to be sure that we always had a probe below the ship's waterline. Comparisons with other contemporary SST measurements demonstrated the high reliability and consistency achieved with these hull SST sensors.

With the advent of satellite tracked and satellite reporting drifting and moored buoys in the mid-1970's we developed a new source of bulk SST measurements. The drifting buoys were outfitted with thermistors that measured in various locations depending on the buoy. The earliest large metal buoy hulls used a thermistor in contact with the buoy hull just below the mean waterline. Later smaller spherical buoys had thermistors protruding from the underside of the buoy's hull. Other small spar buoys had a thermistor somewhere on the lower part of the hull that extended into the water. It should be noted that in addition to this lack of uniformity in terms of the placement of the thermistors there has been little or no coordination in terms of the use of a common type of thermistor or other temperature sensor. In fact most of the SST sensors are not calibrated once they are mounted in the buoy hull. Added to these uncertainties is the fact that the smaller buoys become part of the surface wave field and likely spend a good portion of their time at depths from 1 to 5 m below the surface. The larger buoys will oscillate up and down as well but likely have smaller amplitudes in depth variation than do the smaller buoys.

Moored buoys are also instrumented with thermistors to measure not only the near surface SSTs but also to resolve the profile of the upper ocean using thermistors attached to the mooring bridal. Some moored buoys have a separate float to support a thermistor chain to resolve the temperature profile. Although it is surprising, Emery et al (2000) have shown that the moored buoys measure essentially the same thing as the drifting buoys so the moored and drifting buoy SSTs can be combined. Also they found that the buoy SSTs were a lot less noisy than any of the ship SST measurements. They have evolved to be the standard data set used to calculate the algorithm coefficients in today's calculation of SST from satellite infrared radiances.

All of these ship and buoy SSTs are estimates of some type of bulk SST, which does not actually represent the temperature at the surface of the ocean. This bulk SST is of significant historical importance since it has been used in the formulation of all of our so-called "bulk" air-sea heat flux formulae and because it supplies an estimate of the local heat content. Numerical models today require the input of some form of bulk SST for their computation in spite of the fact that it is the skin SST that is in contact with and interacts with the overlying atmosphere. Some people think that the difference between the skin and bulk SSTs is a constant to account for the cooler skin temperatures. This is not the case as the skin SST is closely coupled to the atmosphere-ocean exchanges of heat and momentum making the bulk-skin SST difference a quantity that varies with fairly short time and space scales depending on the prevailing atmospheric conditions (wind speed and air-sea heat flux).

Satellite SST Algorithms

The computation of SST from infrared satellite data started in the mid-1970's using the primary instrument called the Scanning Radiometer (SR) on NOAA's polar orbiting weather satellites. On the same spacecraft the Very High Resolution Radiometer (VHRR), however, had a 1 km spatial resolution which was much better than the 8 km spatial resolution of the SR. Efforts to fully automate the computation of SST from the satellite radiances produced some very artificial looking SST fields. During this period the only in situ SST data were from merchant ships and these were initially used for comparison with the satellite SSTs. Later it was decided to regress the satellite radiances against the ship data to produce an SST consistent with the traditional bulk SST observations. As a result, the satellite infrared radiances were forced by regression to construct an algorithm to approximate the bulk SST that had been measured by the ships.

The first (1978) Advanced VHRR (AVHRR) had 4 channels that included a mid-range infrared (3.7 μm in channel 3) with a single thermal infrared channel at 11 μm. The 1-km resolution of the AVHRR and its improved radiometric fidelity promised some improvements in the computation of SST. Following earlier practice the SST algorithm coefficients were found by regression of the satellite radiances against a selection of ship SST data. About this time there was significant activity in the deployment and operation of satellite-tracked drifting buoys designed to give a Lagrangian description of the current field. As the number of drifting buoys increased in the world's ocean it became apparent that SSTs collected from these buoys constituted a more accurate and less noisy set of data to use for the computation of the satellite SST algorithm coefficients. Thus, buoy SSTs became the exclusive reference data for the computation of satellite SST algorithm coefficients.

It was realized that atmospheric water vapor markedly attenuates the 11 μm infrared signal from the sea surface and that some way of correcting the AVHRR data for this effect was needed. It was decided to modify the AVHRR to add another thermal infrared channel (~12 μm) that was affected more strongly by atmospheric water vapor attenuation. The difference between the 11 and 12 μm thermal infrared channels would provide an estimate of the water vapor content in the atmosphere which could be corrected for in the SST algorithm. Called the "split-window" formulation this approach has been used in the majority of the SST algorithms developed since this time (McMillin and Crossby, 1984).

Recently the increased use of SST measurements from large moored buoys has led to their being included in the infrared regression procedure. As noted above moored buoy SSTs are treated exactly like drifting buoy SSTs which recent study (Emery et al., 2000) has shown to be a reasonable assumption (moored and drifting buoy SSTs are similar). While drifting buoys caught up in the near-surface wave field measure their "SST" somewhere between 0.5 and 2 m depth, the moored buoys often mix and submerge the waters near the surface thereby measuring some average of data from about 1 - 2 m depth.

There are a number of limitations and problems with this algorithm coefficient estimation procedure. This regression calculation inherently assumes that the buoy data are "perfect" and without error. In their examination of a couple years of buoy data Emery et al. (2000) found that buoy SSTs have typical standard deviations of about 0.3 to 0.5 °C. Added to this error is the fact that the sampling coverage supplied by the drifting buoys is dictated by the need to observe a particular circulation feature (in other words the buoys are not deployed for the purpose of satellite sensor calibration/validation). As a consequence the geographic coverage of these buoy data are less than ideal for the validation of satellite SSTs which compromises the resultant satellite SST computations.

The computation of bulk SSTs from infrared satellite data has progressed relatively little in the past decade. It should be noted that the specific SST algorithm has been changed frequently in an attempt to improve the atmospheric corrections of the IR data. The fundamental approach was the Multi-Channel SST (MCSST; McClain et al. 1985) which was the procedure first associated with the split-window technique. Many changes to the SST algorithm coefficients have occurred over time and often the corrections involved systematic comparisons of the normal MCSST with atmospheric corrections due to high aerosol loading due to volcanic eruptions.

To improve the MCSST scientists at NOAA/NESDIS decided to alter the regression formulation to include non-linear effects (Walton et al., 1990, Walton et al., 1998). The first official alternative algorithm was called the Cross Product SST (CPSST) to emphasize the nonlinear part of the computation. The assumption was that the CPSST would do a better job of compensating for atmospheric moisture (water vapor) especially in regions with high moisture content. Some very modest changes were later made to the CPSST and the algorithm is now known as the non-linear or NLSST. All of these algorithms depend on the split thermal infrared window to correct for atmospheric water vapor. A study by Emery et al (1994) indicated that the AVHRR channel 4, 5 difference is a good indicator of atmospheric moisture in the water laden tropical atmosphere. At mid and high latitudes this 4 - 5 brightness temperature difference is not a function of atmospheric moisture as indicated by a lack of correlation between this temperature difference and the integrated column water vapor content computed from the microwave sensor of the Special Sensor Microwave Imager (SSM/I).

Today the NLSST is the operational algorithm used by the Naval Oceanographic Office (NAVO) and coefficients are still derived from a regression against buoy SSTs (drifting + moored; May et al., 1998). SST validation is carried out using comparisons with independent drifting and moored buoy SST data. It is not clear what direction will be taken at NAVO regarding the evolution of this SST product. At present the emphasis is on the regular computation and delivery of this SST product as a mixture of surface skin (IR satellite) and in situ bulk SSTs.

Skin SST

In the 1960's there were a number of studies that focused on the skin of the ocean in an effort to understand the processes that dictated the SST. Studies (Ewing and McAlister, 1960; Clauss et al., 1970) demonstrated the existence of the skin layer and also showed how a breaking wave momentarily destroys the skin layer. They also found that this layer will reestablish itself within 9-12 seconds after its destruction. More recent infrared camera measurements have shown the skin layer to reestablish itself within less than a second after a breaking wave destroys it (Jessup et al., 1997a) suggesting that the skin layer is almost always present. This restoration of the skin layer depends on the net heat flux and intensity of the background turbulence. Other observational studies (Grassl and Hinzpeter, 1975; Grassl, 1976; Paulson and Simpson, 1981; Katsaros, 1980; Robinson, et al., 1984; Schluessel et al., 1987; Schluessel et al., 1990; Wick et al., 1996; Kent et al., 1996; Donlon and Robinson, 1997) have clearly demonstrated the existence of this thin skin layer at the ocean's surface. Overall bulk-skin SST differences have mean values of about 0.3°C with an RMS variability of up to 0.4°C. The instantaneous value is dependent on the heating/cooling and surface wind conditions.

This skin layer is the molecular boundary between a turbulent ocean and a turbulent atmosphere. This molecular layer is necessary for the transfer of heat, momentum and other properties, between the ocean and the atmosphere. The thermodynamics of this layer also determines the flux of gases such as carbon dioxide between the sea and the atmosphere. Thus, any research into the nature of the air-sea flux of heat, momentum and gases must account for the skin SST while all of the traditional SST estimation has concentrated on the “bulk SST” which is that SST sampled near the surface but not at the surface.

Parametric Model of the Bulk - Skin SST difference (ΔT)

An extensive amount of work has gone into attempting to model and predict the temperature difference across the skin layer. The original models treated the skin layer as a molecular boundary layer where the heat transfer occurs by molecular conduction. The key element of these type models is estimating the thickness of the skin layer. Saunders (1967) developed the first such model. He used dimensional analysis to estimate the skin layer thickness and determined that the temperature difference could be determined from the net heat flux and the wind-driven friction velocity. An alternative approach that has been considered more recently assumes that the temperature difference is controlled by the process of surface renewal (Dankwerts, 1951). In this process, cooler water at the ocean skin is periodically renewed with the warmer bulk water from below. The key parameter governing the magnitude of the temperature difference in this approach is the time between surface renewal events.

In this section we summarize the physical processes that are included in the models and examine the accuracy of current predictions of ΔT. More detailed reviews of the early efforts to model ΔT were presented by Katsaros (1980) and Robinson et al. (1984). Examples of recent work in modeling ΔT are presented by Wick et al. (1996), Fairall et al. (1996) and Eifler and Donlon (2000).

The physical processes that govern the magnitude of ΔT can vary with the environmental conditions. Three different possible regimes are summarized schematically in Figure 1. These regimes include free convection, forced convection driven by wind stress, and forced convection driven by microscale wave breaking. The figure shows the primary processes affecting ΔT in each regime and the representative conditions in the upper ocean and lower atmosphere.

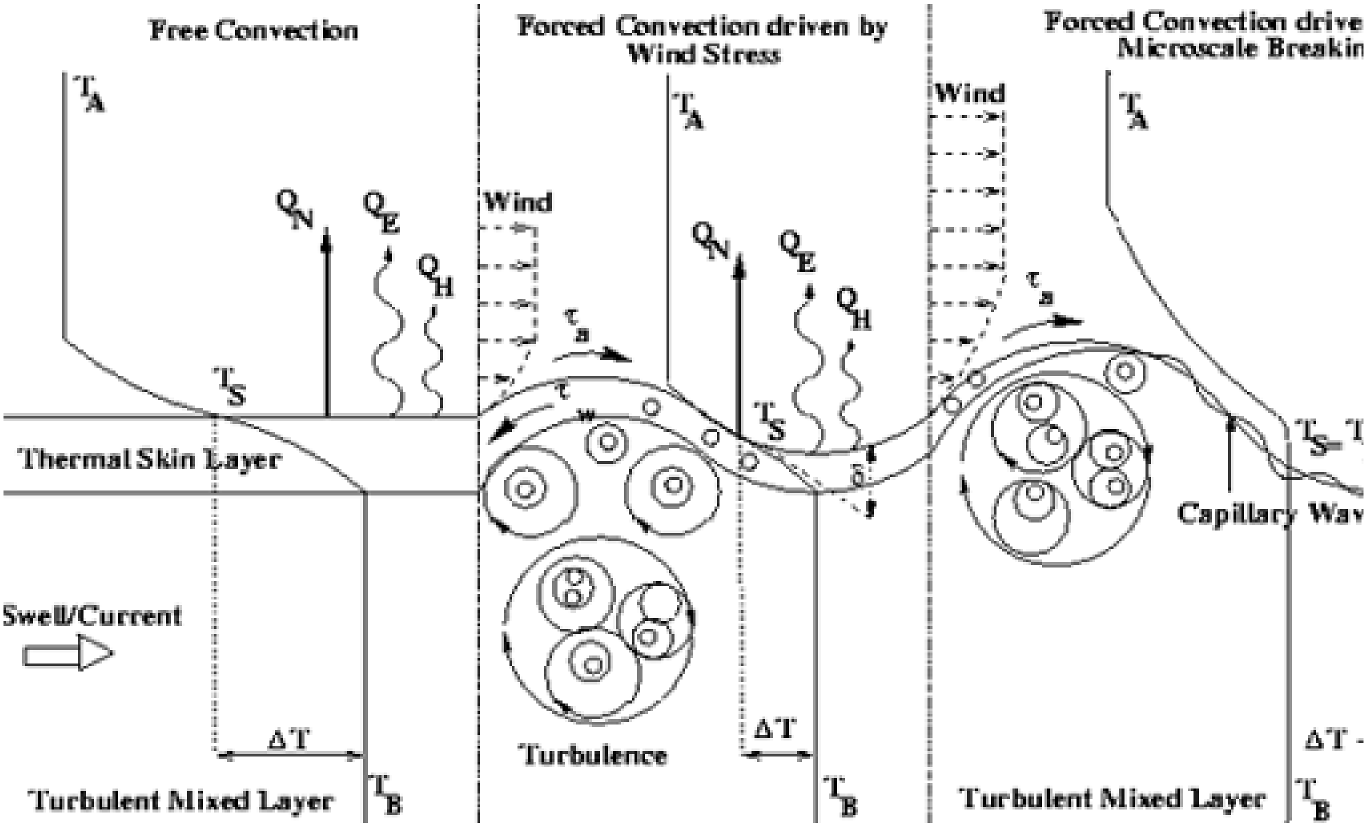

Fig. 1. Behavior of the oceanic skin layer under three primary physical regimes. T_A, T_S, and T_B represent, respectively, the air temperature, oceanic skin temperature, and bulk ocean temperature; Q_E, Q_H, and Q_N represent the latent, sensible, and net radiative heat fluxes, and τ_a and τ_w represent the surface stress in the air and the water.

In the free convection regime which dominates at very low wind speeds, the surface renewal and hence ΔT are governed by static stability. The magnitude of ΔT is controlled primarily by the net heat flux. The skin layer continues to cool as a result of the heat flux until it becomes unstable, sinks, and is replenished by warmer bulk water. In the wind stress driven forced convection regime, ΔT is regulated both by the net heat flux and the wind stress on the surface. The action of the wind on the surface contributes to the renewal of the skin layer. Below the skin layer turbulence acts to create a near isothermal temperature profile but the magnitude of the turbulence is damped as the skin layer is approached. This is the regime originally considered by Saunders (1967) and forms the basis of most oceanographer's approach to the ocean's upper layer.

In the third regime, the surface renewal is dominated by the microscale breaking of very short capillary waves. Microscale breaking refers to breaking without any accompanying air entrainment and microscale breaking waves commonly consist of a bore-like crest accompanied by parasitic capillary waves distributed along the forward face (Jessup et al., 1997b). Recent observations using infrared imaging of the sea surface have demonstrated that microscale breaking can have a significant impact on the temperature of the skin layer (Jessup et al., 1997b; Zappa, 1999). During the breaking, the skin layer is disrupted for a brief interval and the frequency of the renewal may be largely independent of the wind stress and heat flux. The rate of recovery of the skin layer following the breaking will be a function of the heat flux. With frequent enough breaking, the relative effect of the recovery rate on ΔT can be small and the mean ΔT can appear independent of the heat flux and wind speed. An additional regime can also be considered where larger-scale wave breaking with air entrainment disrupts and mixes the skin layer for longer periods of time.

It is important to emphasize that these regimes are not mutually exclusive. The processes associated with each regime can coexist and work simultaneously to regulate ΔT. In fact, recent data demonstrated that a renewal time scale associated with free convection provides a strong limit on the magnitude of ΔT under all conditions. This suggests that while other processes may cause more frequent renewal of the skin layer, cooling of the layer is ultimately limited by static stability.

The ability of several recent models to estimate ΔT is summarized in Figure 2. The models tested include the original model of Saunders (1967), one developed by Soloviev and Schlüssel (1994), one developed by Fairall et al. (1996), and one by Wick et al. (1996). The Fairall et al. model is an enhanced molecular conduction type model that blends the effects of free convection and wind-driven forced convection. The Soloviev and Schlüssel model incorporates a free convection and large-scale breaking regime as well as a moderate wind speed regime that indirectly incorporates smaller-scale wave breaking. Finally, the Wick et al. model combines free convection with a different approach to shear-driven forced convection. Interest in the impact of microscale breaking is only recent and the effects are currently being integrated into new models.

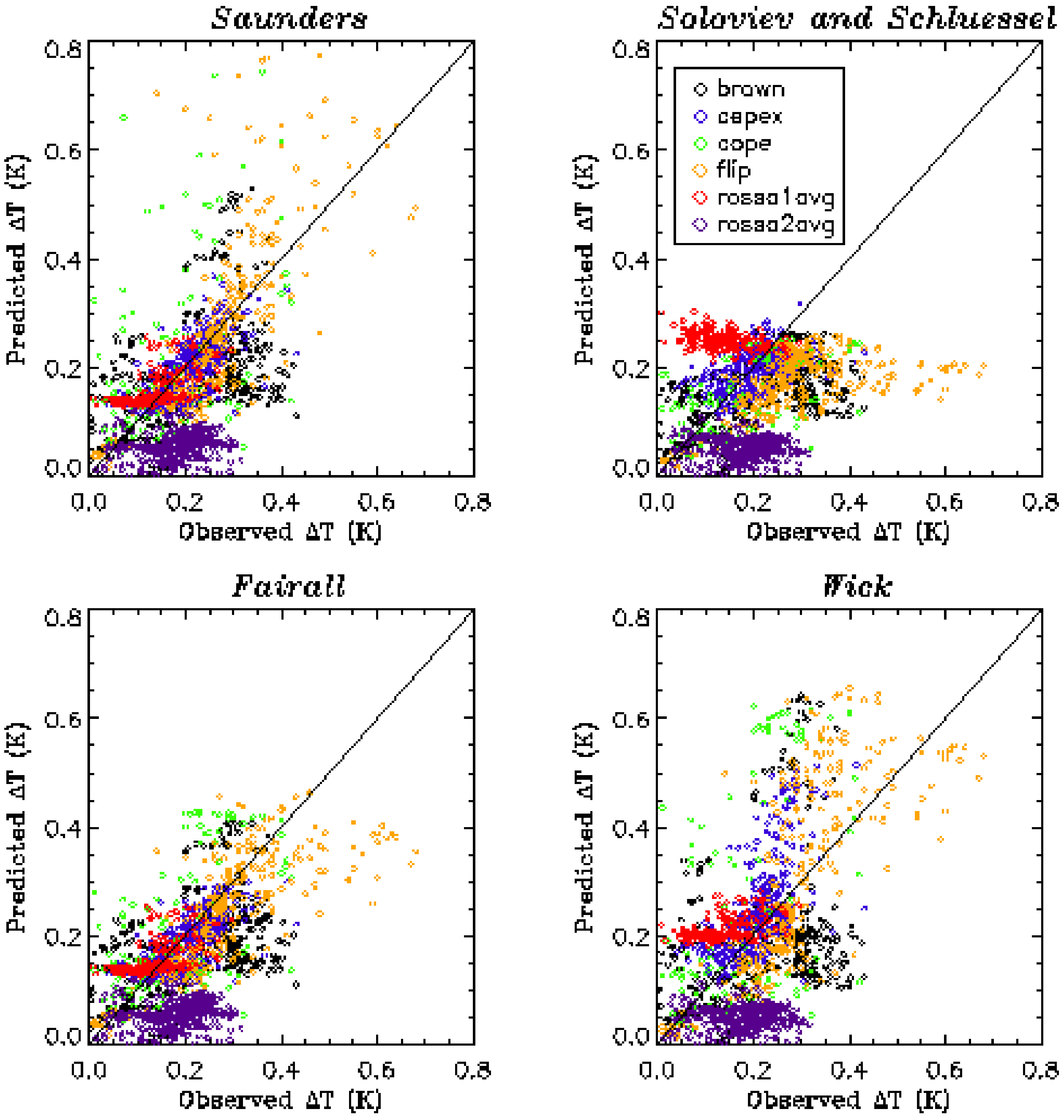

Fig. 2. Observed ΔT against model predicted ΔT for 4 different models and 7 research cruises.

Each model was evaluated using measurements collected during five separate experiments. The different experiments are each represented by different color symbols in Figure 2. The data marked "brown" were collected aboard the R/V Ronald Brown during the Gas Exchange (GASEX) Experiment in the North Atlantic in the spring of 1998. The "cepex" data were obtained during the Central Equatorial Pacific Experiment in 1993. Both the "flip" and "cope" data were collected by Dr. Andrew Jessup of the Applied Physics Laboratory at the University of Washington during experiments on the research platform Flip. The "flip" experiment was conducted off the coast of southern California in 1992 while the Coastal Oceans Probing Experiment (COPE) was performed off the Oregon coast in 1995. Dr. Craig Donlon in the South Atlantic during the Atlantic Meridional Transect collected the data labeled "rossa" in 1996. Overall, the measurements represent a wide range of conditions and seasons.

The results demonstrate that current models can accurately reproduce the overall tendency of ΔT. The agreement is especially interesting given the diverse range of conditions under which the data were collected. Nonetheless, there is still a large amount of variability that the models fail to reproduce. This variability would have a significant impact on any effort to explicitly incorporate modeled ΔT estimates into satellite SST measurements. The errors associated with ΔT estimates at individual satellite pixels would be expected to be even larger due to the lack of accurate flux measurements as available during research cruises. A recent effort to explicitly account for the skin layer in bulk SST measurements from the GOES satellites showed that the same accuracy was obtained whether the skin layer was included or the satellite measurements were directly regressed against in situ bulk SST measurements (Wick et al., 2000). Improvements obtained by including the appropriate physical processes were countered by uncertainties in the models and forcing data.

The best approach to obtaining improved SST accuracy from satellite measurements may be to produce a satellite skin SST product derived through direct regression against a significant number of in situ skin temperature measurements. Hopefully these in situ skin SST measurements will be made with coincident measurements of traditional bulk SST and the measurements used to further improve our model of ΔT so that in the future we can have a more precisely defined relationship between skin and bulk SST. At present the limited number of research skin SST data are inadequate for the routine computation of such a regression skin SST. The required calibration/validation measurements needed to fulfill this requirement are set out later in our discussion of an SST validation plan.

Computing SST from Satellite Data

As is evident from this discussion we are working towards a more complete understanding of the connection between the skin and bulk SSTs. We now understand the driving forces that cause fluctuations in the skin and bulk temperatures and we are able to parameterize the difference between these two temperatures as functions of net air-sea heat flux and wind speed. Unfortunately these parameterizations require a knowledge of net air-sea heat flux to define ΔT and that is much more difficult to observe than the skin or bulk SSTs themselves. It may be in the future that we can use measured ΔTs to compute the net air-sea heat flux. This will require an accurate knowledge of the ocean wind speed but we have new microwave measurement techniques that can measure ocean wind speed and direction with considerable accuracy. What is needed is an increase in the samples of skin and coincident bulk SSTs along with standard meteorological observations so that we can either further augment our models or confirm them as they stand.

As a result of the presently unresolved complexity of computing ΔT it is not yet possible to directly compute bulk SST from radiative measurements of skin SST. It is possible to use indirect methods to estimate the bulk SST from the skin SST (Wick et al., 2000) but the results do not show a general improvement. Acknowledging these limitations in the definition of the relationship between skin and bulk SSTs we presently advocate the parallel computation of both skin and "bulk SST" (the traditional NOAA/NESDIS NLSST) from infrared satellite data. Both of these algorithms employ a "split-window" technique but the two approaches will use different methods for the calculation of the algorithm coefficients for the individual infrared channels and their difference. The traditional approach to computing "pseudo bulk SST" takes a set of edited global drifting and moored buoy SSTs as a basis set. The corresponding satellite infrared radiances are then regressed on the buoy temperatures to yield channel and split-window coefficients for the "pseudo bulk SST." For the skin temperature the general practice (Schluessel et al., 1987) is to use radiative transfer simulations based on a global set of marine radiosonde profiles and the lowest atmospheric temperature in taken to be the skin SST. In the absence of widespread and continuous in situ measurements of skin SST this is the only approach we can use. This method is quite sensitive to the radiosonde data set employed for these calculations. It is very important to have enough radiosonde profiles to be able to include information on all of the different atmospheric conditions that can occur. We have traditionally used a set of over 300 globally distributed marine (ships or small islands) temperature profiles for our simulations.

If a suite of in situ skin SST measurements were available we could drop this use of atmospheric simulations and regress the satellite radiances against the coincident measurements of skin SST. Initially these data are needed to demonstrate that by having in situ skin SST measurements we can improve the accuracy of satellite infrared SST measurements. Since both the in situ and satellite measurements are infrared measurements of the sea surface the accuracy of the satellite measurements with respect to the in situ observations is expected to improve. The remaining errors will be in the atmospheric correction for the skin SST. Eliminating the bulk – skin SST error will focus the efforts of the SST research community on this atmospheric correction problem. Once it is established that in situ skin SSTs are needed, these skin SST calibration/validation data must be transmitted to shore on a realtime basis to make it possible to compute the skin SST algorithm coefficients for the daily production of global skin and bulk (NLSST) SST fields.

Validation of Skin/Bulk SST

As has been stated earlier there is a basic need to regularly collect in situ skin SST data to make it possible to properly compute skin SST estimates from infrared satellite data. The oceanographic community has ignored the fundamental differences between skin and bulk SST for many years trying to explain away variability found in the buoy calibrated satellite SSTs as atmospheric variability. While it is true that the atmosphere is the source of the greatest errors in the satellite SST measurements, the bulk-skin SST difference represents an error source that can be corrected through the use of in situ skin SST measurements.

There are a number of basic issues that must be considered before any particular SST validation plan becomes meaningful. Some of the most important are:

1. We assume that the satellite infrared radiometers measure radiation emitted from within the skin layer of the ocean. Radiometric measurements of the skin temperature will also be affected by reflected radiation from the sky and clouds. Radiometers used to validate the

satellite estimates must, therefore, measure both the ocean-emitted radiation and the incoming atmospheric radiation to enable compensation for the reflection effects.

2. The infrared radiation emitted by the sea surface is attenuated in the atmosphere by water vapor, aerosols and clouds. We are only interested in radiances in the absence of clouds so we require that the infrared satellite data be filtered to remove cloud-contaminated pixels. We also assume that a "split-window" algorithm formulation at least partially corrects for the present of atmospheric moisture.

3. We must assume that in terms of global statistics, historical radiosonde profiles are representative of the full range of atmospheric conditions.

4. We need routine and automatic in situ measurements of both skin and bulk SST collected under relatively "cloud-free" conditions. We will require that any set of merchant and/or research ships instrumented for SST validation will be sufficient to cover the same full range of oceanic and atmospheric conditions. Following Emery et al. (2000) this means that we cover the global range of latitude in at least both major ocean basins.

5. We need cost-effective, accurate and reliable shipboard skin SST radiometers to be available in the very near future for deployment and operation on ships of opportunity. These radiometers must be capable of measuring the ocean skin SST to within ±0.1°C while working autonomously. Efforts will be made to extend these radiometric SST measurements to moored buoys.

6. It will be necessary to continue to collect bulk SST observations from drifting and moored buoys and it will also be necessary to deploy some buoys specifically for the collection of in situ bulk SST data. These latter buoys will require unique drogue installation to maintain ocean position rather than to couple tightly with the ocean and flow with the current.

The resolution of these issues leads us to the design of an SST calibration/validation system that initially concentrates on ships of opportunity (SOO) which are merchant vessels that make regular and repeated transects of the major oceans. These ships have already collected a considerable amount of oceanographic data in the past. In addition to routine 12 hourly meteorological observations (including bulk SST) many merchant ships have collected upper ocean temperature profiles using expendable bathythermographs (XBTs) along with a variety of meteorological measurements. Most of these data have been transmitted by radio back to a central location so that the data are available in near realtime.

Regular and repeated measurements of skin and bulk SST are a necessary part of any SST validation program since the validation needs to take place on a regular basis with data relayed in near realtime. It is not satisfactory to validate the satellite infrared sensor once and then leave it forever. The satellite sensors drift over time adding errors to the residual atmospheric effects creating random errors in the SST estimation. A portion of the sensor error is also provided by errors in the onboard sensor calibration process. Continuous global measurements of both skin and bulk SST are required to monitor any changes in the sensor or measurement accuracy.

Availability of both in situ skin and bulk data will enable an assessment of the benefits and accuracy improvements possible through production of a satellite skin SST product. Satellite skin SST algorithms will be derived through regression to the in situ skin measurements. The accuracy of these products as determined through comparison with independent in situ skin measurements can then be compared with that of traditional bulk SST products and simulation-based skin algorithms. We can also use the many new coincident bulk and skin SSTs with estimates of the heat flux derived from the bulk formulae to further explore the relationships between the bulk and skin temperature difference and the driving forces of the wind stress and net air-sea heat flux.

Specific Validation Plan

In situ infrared radiometer systems

For direct validation of satellite skin SST products, radiometers of similar spectral wavebands are required to determine in situ skin SST contemporaneously with the satellite instrument. The widespread deployment of contemporary in situ research instrumentation such as the multi-channel infrared sea truth radiometric calibrator (MISTRC, see Saurez, 1997), the scanning infrared sea surface temperature radiometer (SISTeR see Donlon and Nightingale, 1999), or the Marine Atmospheric Emitted Radiance Interferometer (M-AERI, see Smith et al, 1996) from dedicated research vessels will continue to provide a limited source of high precision in situ skin SST measurements. Both dedicated and opportunistic research cruises have produced a wealth of in situ skin SST data but the number of successful satellite validation instances has been disappointingly small. This is due to the availability and difficulty of developing, maintaining and deploying research instrumentation on suitable ships which, themselves need to be positioned in clear sky conditions and within the satellite swath. Directly addressing this limitation is the main goal of this proposal: we identify a need to shift from a small number of highly sophisticated and dedicated scientific measurement campaigns to the use of accurate, less sophisticated (and therefore lower cost) autonomous in situ radiometer systems that can be widely deployed from opportunistic platforms and ships.

Requirements of in situ skin SST measurements for satellite SST validation

An accuracy of ±0.1 K root-mean-square error is considered the minimum requirement for an in situ radiometer measurement to be eligible for satellite SST validation work. Due to the complexity of making such an accurate in situ skin SST measurement in the harsh marine environment, the following design features are considered mandatory for operational radiometer systems:

i. Self-calibration of the detector output for all measurement ensembles to fully account for instrument drift.
ii. Appropriate views of both the sea surface and the sky are required to explicitly account for reflection of sky radiance at the sea surface.

iii. An appropriate filter response is required that is broadly consistent with the satellite instruments (typically 8-12μm).
iv. Adequate protection of all instrument fore-optics in all conditions to prevent degradation of calibration sources or instrument fore-optics.

In addition, an accurate knowledge of the sea water emissivity (e.g. Bertie and Lan, 1996) for the specific instrument spectral response and deployment view geometry is also required to attain a skin SST accuracy of ±0.1 K (Donlon and Nightingale, 1999).

Current state of the art autonomous in situ skin SST systems

There has been considerable development of low-cost infrared radiometer systems suitable for extended autonomous operation focussed on the use of "off the shelf" solid state radiometers (e.g., Donlon et al., 1998). Based on this early work, two new autonomous instruments have been developed that will commence pre-operational trials using commercial shipping in early 2001. The calibrated infrared in situ measurement system (CIRIMS, Jessup, 1999) is a broad-band (7-16μm) self-calibrating radiometer that is completely sealed from the environment. This instrument has been specifically developed at the University of Washington for validation of the MODIS SST algorithm. All measurements are made through a renewable infrared window providing protection for the instrument calibration and fore-optics system. An alternative design has been adopted by the Infrared Autonomous SST radiometer (ISAR, Donlon et al, 2000) developed for the European ENVISAT AATSR validation program. This design uses an accurate optical rain sensor to trigger a shutter mechanism that completely seals the radiometer fore-optics and calibration black body cavities during bad weather. The ISAR has a spectral bandwidth of 9.6-11.5μm and has the additional capability to report data in near real time via satellite link. The ISAR can also connect several other sensors to the system (e.g. wind speed, bulk SST sensor) and provide a complete in situ SST validation measurement system. Both of these instruments adopt two blackbody calibration cavities and obtain views of both the sea surface and atmosphere each measurement ensemble.

Traceability of calibration

The scientific merit and usefulness of in situ skin SST data is critically dependent on the specific instrumentation and the deployment method used. For each particular instrument, attention must be given to the traceable end-to-end calibration of all internal calibration methods and temperature sensors. At a minimum, independent calibration against a tracable calibration reference blackbody (such as the CASOTS cavity specifically designed for this purpose (Donlon et al, 1999)) before and after each deployment is required. Additionally, radiometer inter-calibration exercises are required to inter-compare radiometer systems both in the laboratory and in the field. Two such workshops have taken place to date (Donlon et al, 1998; Kannenburg 1998) that have considerably furthered the understanding and requirements for SOO in situ radiometer systems.

Ships of Opportunity (Merchant Ships)

The primary source of continuous in situ validation data for satellite skin and bulk SSTs will be radiometric and hull measurements of skin and bulk SST made from merchant vessels.

Emery et al., (1997) used such ships to collect hull bulk SSTs and comparisons with other sources of coincident bulk SSTs were excellent suggesting that this method is far superior to using ship injection temperatures and is comparable to drifting buoy bulk SSTs. These bulk SST measurements were accurate to ±0.1°C. A map of all the merchant ship tracks in 1996 is shown here in Fig. 3, which clearly shows that the ship traffic is primarily in the Northern Hemisphere.

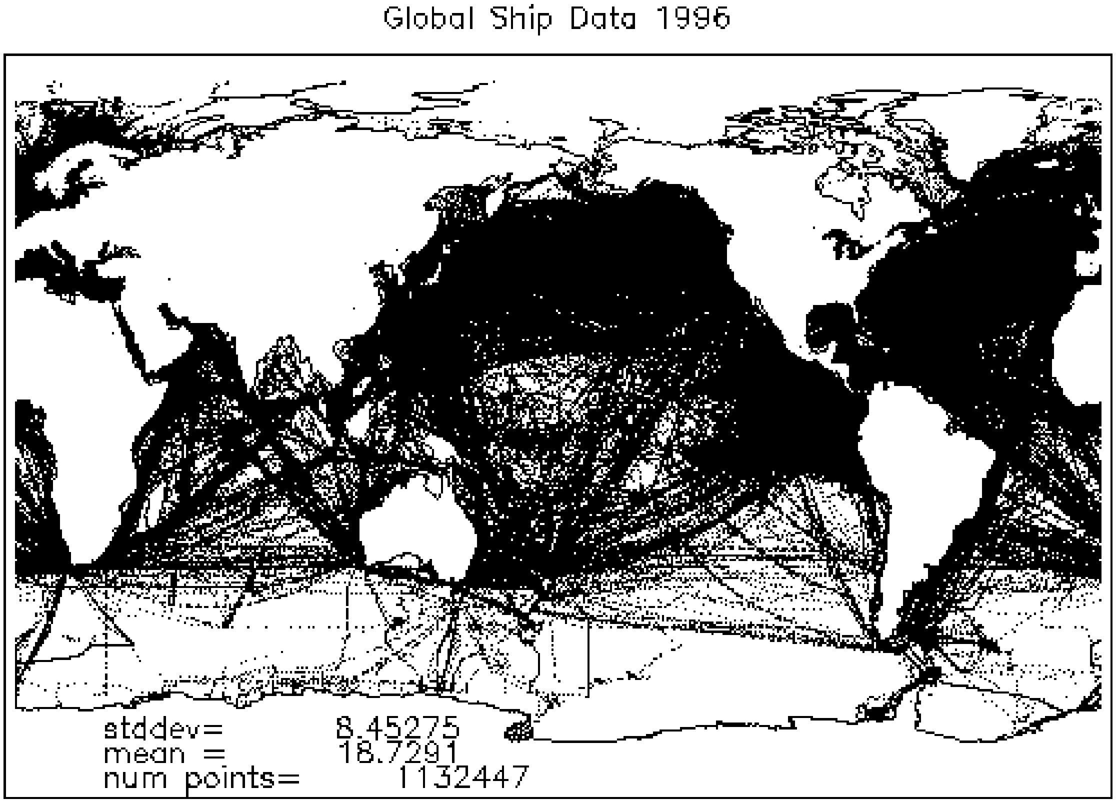

Fig. 3 Ship tracks for 1996.

Fortunately there are some regular routes between places such as Seattle to Auckland and Tokyo to Valpariso. These long tracks offer the opportunity of measuring conditions over fairly large distances and they cover a considerable range in latitude. The only area unsampled is the southern sub-polar region which is only crossed in the austral summer by Antarctic resupply ships. These ships are both research vessels as well as supply ships and can be employed to collect measurements en route to the Antarctic.

Once we have been able to demonstrate that these measurements can be collected reliably from merchant vessels, we propose to instrument approximately 20 ships with infrared skin SST radiometers, hull temperature bulk SST sensors and an improved suite of meteorological measurement systems. A "strawman" selection of merchant ship routes is shown here in Fig. 4 based on the abundant ship tracks in Fig. 3. To provide realtime data we should take advantage of the best possible communication options (cellular phone, satellite communications, ARGOS, etc.) at the time to insure the timely, accurate and safe transmission of data from these instruments. The selected ships will all traverse long nearly north-south routes in the ocean to provide the greatest possible coverage and resolution. Thus, even if the central portions of the ocean are not completely covered, the errors will be about the same if the north-south coverage is adequate (Emery et al., 2000). The important result is that we have a global system continuously reporting simultaneous skin and bulk SSTs along with the corresponding meteorological data to compute the heat flux.

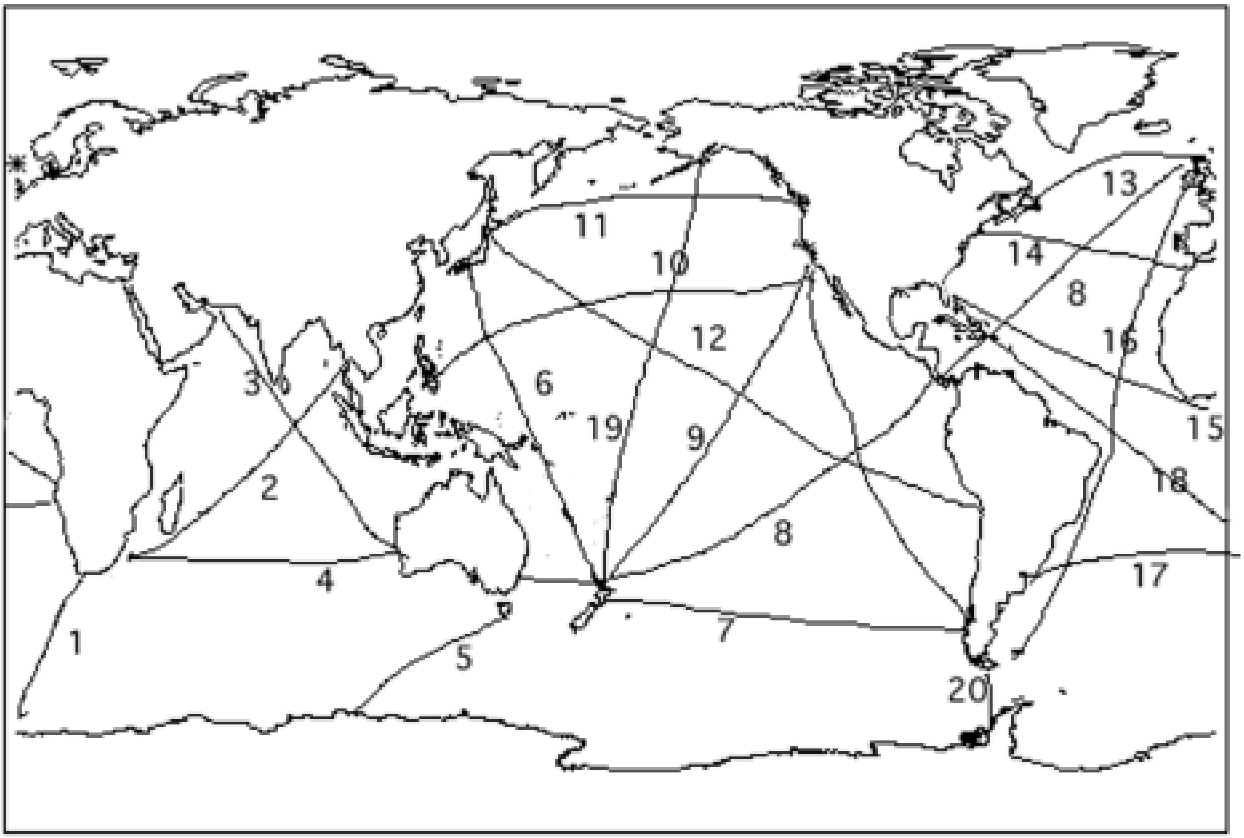

Fig. 4 Selection of 20 global merchant ship routes for infrared SST calibration/validation.

Moored and Drifting Buoys

One advantage of installing radiometers on ships of opportunity is that it is possible for a person to check on the radiometer and its operation each day. The likelihood of sea spray contaminating the radiometer optics is quite high even if the system has been designed to avoid this contamination. This is the major limiting factor in the autonomous operation of infrared radiometers from moored buoys. Once skin SST calibration radiometers have been developed that can perform under all sea conditions in an automated fashion we want to extend our SST radiometer sampling network to include moored buoys which already measure the bulk SST.

The advantage of moored buoys is that they can operate in ocean areas that are not crossed by merchant ship traffic. It is unlikely that an SST validation program can afford to maintain a large number of moored buoys but it is possible for such a validation program to instrument buoys operated for other programs. Our goal is to cover as large a range of latitudes as possible and we advocate cooperation with moored buoy programs that offer latitude coverage not available from ships of opportunity. At the moment there are the TOGA TAO buoys in the equatorial Pacific along with moored National Data Buoy Center (NDBC) buoys along the east and west coasts of the US. A number of other buoy programs are planned and it is hoped that in the near future we will have moored buoys in the central mid-latitude Pacific as well as the Atlantic.

The mainstay of the present NOAA/NESDIS and NAVO SST validation/calibration programs producing routine (bulk) SST are the drifting buoy SSTs. It should be pointed out that the majority of these buoys were not deployed to provide calibration SST for satellite data but rather primarily to map ocean currents or make some other relevant measurement. A consequence of this fact is that the geographical coverage is frequently not ideal for SST calibration/validation. An example is presented here in Fig. 5, which is the global distribution of buoy bulk SST measurements in October, 1990.

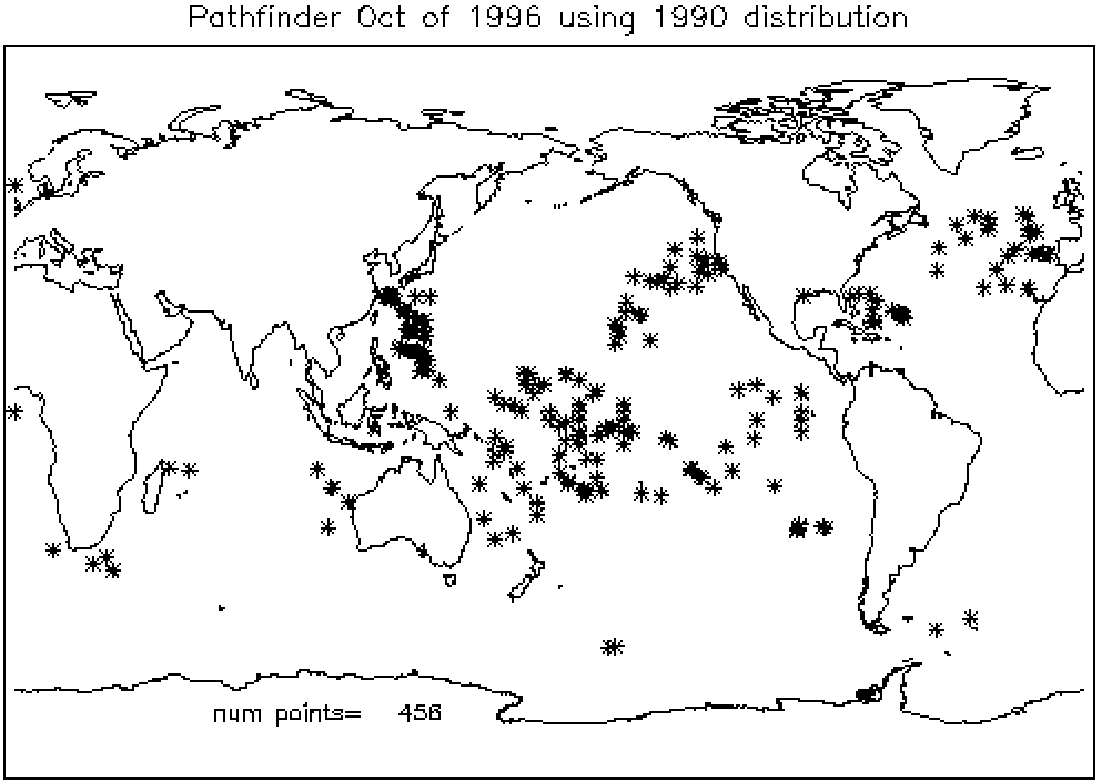

Fig. 5 The October, 1990 distribution of buoy bulk SST measurements.

As can be easily seen there are almost no buoy SSTs in the Indian Ocean, or in the South Atlantic. The central Pacific is covered well but the north and south extremes are unsampled. Fortunately the overall, global north-south coverage is not too bad resulting in SST coefficients that were just slightly different than those from a year with better coverage (Emery et al., 2000). Still if one were to design a program for SST validation it would not recommend this type of geographic coverage. Thus, an important recommendation is that we need to have a drifting buoy program strictly for the validation/calibration of satellite infrared SSTs. In such a program the buoy deployment would be selected to optimize the global and temporal coverage. Also, these buoys would be instrumented not to follow ocean currents but rather to move horizontally as little as possible (perhaps with deep drogues to follow the weaker deep currents). In this application we can take advantage of the known mean current patterns to optimize the spatial distribution of the buoy SSTs. Combined with the expanding network of moored buoys and the ships of opportunity these drifters should enable us to obtain an excellent space/time sample of the bulk SST. It will be very interesting to compare the ship hull bulk SSTs with these buoy SSTs.

The selection of regions for the deployment of buoys in this validation scheme will be based on the coverage of the merchant vessels given for example in Fig. 4. From this map of sampling routes it is clear that we will need additional measurement in the Southern Ocean (south of 30 °S) and in the Indian Ocean. The other areas will be well sampled by the merchant ship array of sensors. Since the Antarctic Circumpolar Current (ACC) will move buoys in the Southern ocean rapidly to the east, we should plan to deploy buoys that will continue to circuit around the Antarctic continent during their period of operation. There is a need for increased sampling of atmospheric pressure in this region and it should be possible to share our buoy measurements with the weather services of the world.

Validation Plan Summary

The validation plan is designed not to carry out a one-time comparison data set to validate the satellite sensors but instead to provide the repeat measurements needed to sustain the calibration/validation of the infrared satellite measurements. It should be pointed out that in the future we may be able to use our in situ skin SSTs for validation/calibration of passive microwave SSTs in addition to the IR radiances we have been discussing. To cover the important meridional changes in the environment, the validation plan is made up of ship of opportunity, moored and drifting buoy measurements. For the ships and moored buoys, radiometers will collect skin SST measurements coincident with the in situ bulk SST observations. Drifting buoys will measure bulk SST in regions not occupied by ships or moored buoys. Together these measurements will provide us with continuous global measurements of both skin and bulk SST to validate and/or modify the satellite measurements. With these validation measurements it should be possible to achieve an SST accuracy of 0.1 – 0.3 °C for both skin and bulk SST. Furthermore, this body of skin and bulk SST measurements should dramatically improve our knowledge of the air-sea heat flux and its variability in time and space.

Acknowledgements

The authors would like to thank Andrew Jessup for the use of the FLIP and COPE data, Mike Reynolds for his efforts on the Brown cruise and Craig Donlon for all of his efforts on the ROSSA cruises. The National Aeronautics and Space Administration (NASA) under their Earth System Enterprise program, Jim Dodge and Eric Lindstrom provided program managers funding for many of these experiments. Earlier funding was supplied by the German Research Agency (DFG).

References

Bertie J E and Z D Lan (1996) Infrared intensities of liquids: the intensity of the OH stretching band revisited, and the best current values of the optical constants of H_2O (l) at 25°C between 15,000 and 1 cm^{-1}. *App. Spectroscopy* 50: 1047-1057.

Clauss, E., H. Hinzpeter, J. Mueller-Glewe, 1970: Messungen der Temperaturstruktur im Wasser and der Grenzflaeche Ozean-Atmosphaere. Meteor-Forschungs-Ergebnisse, Reihe B, No. 5, 90-94.

Dankwerts, P. V., 1951: Significance of liquid-film coefficients in gas ablorption. *Ind. Eng. Chem.*, **43**, 1460-1467.

Donlon, C. J., and I. S. Robinson, 1997: Observations of the oceanic thermal skin in the Atlantic Ocean. *J. Geophys Res*., 102, 18 525- 18 696.

Donlon C J, S J Keogh, D J Baldwin, I S Robinson, I Ridley, T Sheasby, I J Barton, E F Bradley, T J Nightingale and W Emery, (1998) Solid state measurements of sea surface skin temperature. *J. Atmos. Oceanic Tech.* 15: 775-787.

Donlon C J and T J Nightingale (2000) The effect of atmospheric radiance errors in radiometric sea surface skin temperature measurements *Appl. Optics* 39:2397-2392.

Eifler, W. and C. J. Donlon, 2000: Modelling the thermal surface signature of breaking waves, submitted to *J. Geophys. Res*.

Emery, W. J., Y. Yu, G. A. Wick, P. Schlüssel, and R. W. Reynolds, 1994: Correcting infrared satellite estimates of sea surface temperature for atmospheric water vapor attenuation. *J. Geophys. Res*., 99, 5219-5236.

Emery, W.J. and K. Cherkauer, B Shannon and R.W. Reynolds 1997: Hull mounted bulk sea surface temperature measurements from volunteer observing ships. *J. Atm. Oceanic Tech*. 14 , 1237-1251.

Emery, WJ, DJ Baldwin, P. Schluessel, RE Reynolds, 2000: Accuracy of In Situ Sea Surface Temperatures Used to Calibrate Infrared Satellite Measurements, submitted to *J. Geophys. Res*.

Ewing, G. and E.D. McAlister, 1960: On the thermal boundary layer of the ocean. *Science*, 131, 1374-1376.

Fairall, CW, EF Bradley, JS Godfrey, GA Wick, JB Edson and GS Young, 1996: Cool skin and warm layer effects on sea surface temperature. *J. Geophys. Res*., 101, 1295-1308.

Grassl, H. and H. Hinzpeter, 1975: The cool skin of the ocean. GATE-Rpt. 14, 1, WMO/CSU, Geneva, 229-236.

Grassl, H., 1976: The dependence of the measured cool skin of the ocean on wind stress and total heat flux. *Boundary Layer Met*., 10, 465-474.

Jessup, A. T., C. J. Zappa, M. R. Loewen, and V. Hesany, 1997a: Infrared remote sensing of breaking waves, *Nature*, 385, 52-55.

Jessup, A. T., C. J. Zappa, and H. Yeh, 1997b: Defining and quantifying microscale wave breaking with infrared imagery. *J. Geophys. Res*., **102**, 23,145-23,153.

Katsaros, K.B., 1980: The aqueous thermal boundary layer. *Bound.-Layer Meteor*., 18, 107-127.

Kent, E. C., T. N. Forrester, and P. K. Taylor, 1996: Comparison of oceanic skin effect parameterizations using shipborne radiometer data. *J. Geophys. Res*., 101, 16 649-16 666.

May, D. A., M. M. Parmeter, D. S. Olszewski, and B. D. McKenzie, 1998: Operational processing of satellite sea surface temperature retrievals at the Naval Oceanographic Office. *Bull. Amer. Meteor. Soc*., **79** 397-407.

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

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

Paulson, C. A., J. J. Simpson, 1981: The temperature difference across the cool skin of the ocean. *J. Geophys. Res*., 86, 11,044 - 11,054.

Robinson, IS, NC Wells and H. Charnock, 1984: The sea surface thermal boundary layer and its relevance to the measurements of sea surface temperature by airborne and spaceborne radiometers. *Int J. Rem. Sens*. 5, 19 - 45

Saunders, P., 1967; The temperature at the ocean-air interface. *J. Atmos. Sci*., 24, 269-273.

Saur , J.F.T., 1963: A study of the equality of sea water temperatures reported in logs of ship's weather observations, J. App. Meteorol., 2, 417-425.

Schluessel, P., H.-Y. Shin and W.J. Emery, 1987: comparison of satellite-derived sea surface temperatures with in situ skin measurements. *J. Geophys. Res*., 92, 2859-2874.

Schluessel, P., W.J. Emery, H. Grassl and T. Mammen, 1990; 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.

Soloviev, AV, and P Schlüssel, 1994: Parameterization of the cool skin of the ocean and of the air-ocean gas transfer on the basis of modeling surface renewal. *J. Phys. Oceanogr*., 24, 1339 - 1346.

Walton, CC, EP McClain, and JF Sapper, 1990: Recent changes in satellite-based multi-channel sea surface temperature algorithms, Science and Tech for a New Orleans Decade, MTS 90, Mar. Tech. Soc., Wash. DC, Sept. 1990.

Walton, C. C., W. G. Pichel, J. F. Sapper, and D. A. May, 1998: The development and operational application of nonlinear algorithms for the measurement of sea surface temperatures with the NOAA polar-orbiting environmental satellites. *J. Geophys. Res*., **103**, 27,999-28,012.

Wick, G. A., W. J. Emery, L. H. Kantha, and P. Schlüssel, 1996: The behavior of the bulk-skin temperature difference under varying wind speed and heat flux. *J. Phys. Oceanogr*., **26**, 1969-1988.

Wick, G. A., J. J. Bates, and D. J. Scott, 2000: Satellite and skin layer effects on the accuracy of sea surface temperature measurements from the GOES satellites. submitted to *J. Atmos. Oceanic Technol*.

Zappa, C. J., 1999: Microscale wave breaking and its effect on air-water gas transfer using infrared imagery. Ph.D. Thesis, University of Washington, 202 pp.