The Variability of Sea Surface Temperature and the Impact it has upon Climate Modelling

by

Ian K. Ridley

Department of Physics and Astronomy,
University of Leicester

A thesis presented for the degree of Doctor of Philosophy

April, 2000
ABSTRACT

The Variability of Sea Surface Temperature
And the Impact is has upon Climate Modelling

Ian K. Ridley

A set of in situ meteorological and radiometric measurements is acquired to facilitate the investigation of skin sea surface temperature (SSST) variability. The Tasco THI-500L radiometer produces SSST accurate to ±0.5K when deployed in-shore and calibrated every 72 hours. Radiometer performance is enhanced if the instrument is insulated from, and characterised for, the effect of solar heating, allowing operational deployment to improve the availability of in situ SSST.

Two statistical tests for similarity of populations are evaluated as quantifiers of sea surface temperature variability. Violations of the parametric requirements of the analysis of variance F-Test produce unreliable results. A dependable measure of SSST variability is generated using the non-parametric Mann-Whitney U-Test. Applying this test to imaging radiometer measurements shows that $U$ depends on the variation between radiometric images in the physical factors that govern the ocean-atmosphere heat fluxes. The implications of this dependency for satellite SSST validation and climate modelling are considered. Analysis over metre- to kilometre- scales produces three trends of SSST variability against measurement integration time. No dependency of these trends on meteorological conditions is found.

Applying the U-Test to large-scale ATSR-1 ASST data of the Tropical Pacific Ocean finds less variability than in small-scale SSST. Large-scale SSST are shown to be inhomogeneous with co-incident bulk sea surface temperature (BSST), implying that the use of ATSR-1 SSST in climate models will impact on the model results.

Forcing a model of the Tropical Pacific Ocean separately with SSST and BSST shows a divergence in model output for both seasonal and ENSO SST variability. The BSST output is 0.7K warmer than the SSST output in the central equatorial Pacific during the Spring warming period. This area of divergence extends eastwards during the 1991-2 ENSO event, corresponding to the characteristic easterly movement of warm water.
2.4.4 Calibration strategy
2.4.4.1 Tasco THI-500L radiometer
2.4.4.2 Thermal Infrared Camera Calibration
2.4.5 Other Pre-processing issues
2.4.5.1 Noise on Tasco THI-550L data from the “Yokogowa” logger
2.4.5.2 Sunglint in the Thermal infrared camera data

2.5 Conclusions

3. THE VARIABILITY OF SSST AT SMALL-SCALES

3.1 The need to investigate SSST variability at small-scales

3.2 Retrieval of in situ SSST from the NEC TH1101

3.3 NEC TH1101 scan geometry

3.4 Statistical tests for the variability of sea surface temperature
3.4.1 Defining “variability”
3.4.2 The F-Test for analysis of variance
3.4.2.1 Definition and calculation of the F-Test
3.4.2.2 Effects of non-normality, unequal variances and unequal sample sizes on the F-Test
3.4.2.3 Results of the F-Tests
3.4.2.4 The dependency of the F-Test on measured physical parameters
3.4.3 The Mann-Whitney U-Test
3.4.3.1 Definition and calculation of the U-Test
3.4.3.2 Results of the U-Tests
3.4.3.3 The dependency of U on the measured physical parameters
3.4.4 The effect of the sunglint-detection algorithm on F and U-test results
3.4.5 The form of the dependency of F and U-Tests on observed physical parameters
3.4.6 Is SSST variability governed by the same factors as air-sea heat fluxes and ΔT?
3.4.7 Comparison of F and U-Tests results

3.5 Variability of TIC image Brightness Temperature: implications for satellite SSST validation
3.5.1 Methodology: investigation of standard deviation
3.5.2 Results Analysis

3.6 Conclusions

4. COMPARISON OF LARGE-SCALE TROPICAL PACIFIC SSST AND BSST DATASETS

4.1 Overview

4.2 The Along Track Scanning Radiometer-1 (ATSR-1) instrument
6.1 Conclusions of this study
6.1.1 Introduction 161
6.1.2 Field Measurements 161
6.1.3 Investigations of sea surface temperature variability 163
6.1.4 Ocean model analysis 167

6.2 Suggestions for Further Work 168

7. BIBLIOGRAPHY 170
Acknowledgements

I would like to thank Dr. Sean Lawrence, my supervisor, and Professor David Llewellyn-Jones for their help, suggestions and guidance throughout the entirety of this work. Other debts of gratitude are due to:

Dr. Isabelle Parkes, Will Rathman, Tom Sheasby and Dr. Jon Angell at the Earth Observation Science Group, Department of Physics and Astronomy, University of Leicester.

Professor Ryuzo Yokoyama, Dr. Sumio Tanba, T. Suoma and the rest of the MuBEx '95 group from Iwate University, Japan.

Dr. Tim Nightingale, Albin Zavody and Dr. Chris Mutlow at the Rutherford Appleton Laboratory.

The Aomori Aquaculture Research Centre for their continued support of MuBEx '95, the crew of the \textit{Dai-Ni Misago} and Mr Hosokawa, owner of the \textit{Dai-Ni Misago}, for the use of his vessel.

Dr. Craig Donlon (now at JRC, Ispra, Italy), Duncan Baldwin of Southampton University and Dr. Ian Barton of CSIRO for providing instrumentation and invaluable assistance in acquiring experimental data.

I would like to acknowledge the support of NERC for providing a studentship and travel funds for 3 years of my studies in Leicester. I also wish to acknowledge the financial and technical sponsorship for the MuBEx project provided by NASDA, Rutherford Appleton Laboratory, NERC, the Royal Society, DAIWA Bank, the British Council, the University of Southampton, U.K., the Japan Society for the Promotion of Science and the Japanese Ministry of Education.

Finally my heartfelt thanks go to my parents, Vivienne and Keith, and my wife Sarah for their patience and endless support.
LIST OF FIGURES

Figure 1.1: The Planck emission spectra for a black body at various temperatures .......... 15
Figure 1.2: Composition of the radiometric signal from the sea surface ......................... 16
Figure 1.3: Atmospheric transmission in the visible and infrared .................................. 17
Figure 1.4: Heat flow at the ocean/atmosphere interface leading to a cool skin ............... 24
Figure 1.5: Absorption of solar radiation in clear water (data from Schmidt, 1908) .......... 25
Figure 1.6: Formation and destruction of a diurnal thermocline (Yokoyama et al., 1995) .. 25
Figure 1.7: Difference between observed and predicted $\Delta T$ for models tested by Kent et al. (1996) for wind speeds a) <3 ms$^{-1}$ and b) >3 ms$^{-1}$ .............................................. 30
Figure 2.1: Location of Mutsu Bay Experiment and buoy positions ................................. 39
Figure 2.2: Response function of the Tasco THI-500L radiometer (data courtesy of Tasco Corporation, Japan) ........................................................................................................ 41
Figure 2.3: NEC TH1101 response function (courtesy NEC Corporation, Japan) ............ 41
Figure 2.4: The deployment of instrumentation on the R/V Dai-Ni Misago ....................... 45
Figure 2.5: Emissivities in the 8-12$\mu$m band for various black body designs and radiometer spot radii (Nightingale, 1996) ........................................... 48
Figure 2.6: Spectral emissivity of Nextel 101-C10 (data from Clarke and Larkin, 1985) .. 48
Figure 2.7: The calibration reference black body ......................................................... 49
Figure 2.8: Intercept coefficients from polynomials fitted to the six calibrations of the sea-pointing Tasco against the black body calibration source during MuBEx '95 ............ 50
Figure 2.9: $x$ coefficients from polynomials fitted to the six calibrations of the sea-pointing Tasco against the black body calibration source during MuBEx '95 ............... 51
Figure 2.10: $x^2$ coefficients from polynomials fitted to the six calibrations of the sea-pointing Tasco against the black body calibration source during MuBEx '95 ............... 51
Figure 2.11: Exponential curve fitted to solar flux calibration data .................................. 54
Figure 2.12: Quadratic curve fitted to solar flux calibration data .................................... 55
Figure 2.13: Raw automatically logged Tasco data (black) and manually logged interpolated Tasco data (red) for 22/08/95 observation cruise ........................................ 58
Figure 2.14: Power spectrum of $S^a - S^m$ for 22/08/95 calibration data using interpolated manually logged data to remove signal trend ...................................................... 59
Figure 2.15: Power spectrum of a 7-point running mean of the $S^a - S^m$ difference for 22/08/95 data .................................................................................................................. 61
Figure 2.16: Power spectrum of $S^a - S^m$ minus a 7-point running mean of $S^a - S^m$ for 22/08/95 data .................................................................................................................. 61
Figure 2.17: Power spectrum of $S^a - S^\gamma$ for 22/08/95 calibration data using a 7-point running mean to remove signal trend ............................................................................................................62
Figure 2.18: Distribution of $S^a - S^\gamma$ difference values for 22/08/95 observation data ............ 63
Figure 2.19: Raw automatically logged Tasco data after removal of anomalous noise peaks (black), and manually logged interpolated Tasco data (red) for 22/08/95 observation cruise ........................................................................................................................................................................... 64
Figure 3.1: Along track pixel dimension geometry for the TIC pixel at the centre of the scan line ........................................................................................................................................................................ 73
Figure 3.2: Along track TIC pixel dimension geometry at the end of the scan line as seen from the side (a) and the front (b) ........................................................................................................................................ 74
Figure 3.3: Across track pixel dimension geometry for a TIC pixel at the centre of the scan line .......................................................................................................................................................... 75
Figure 3.4: Across track pixel dimension geometry for the TIC for a pixel at the end of the scan line .......................................................................................................................................................... 75
Figure 3.5: Variation of TIC pixel dimensions with the across-track position of the pixel..... 76
Figure 3.6: Comparison of TIC pixel footprint area for quadrilateral and elliptical- shaped footprints ................................................................................................................................................76
Figure 3.7: F-Test statistic versus Meteorological Parameters ........................................................................................................ 88
Figure 3.8: Fitted relationship for variation in wind speed and $F$ value ........................................... 91
Figure 3.9: Fitted relationship for variation in solar flux and $F$ value ........................................... 91
Figure 3.10: Fitted relationship for variation in relative humidity and $F$ value ................................. 92
Figure 3.11: U-Test statistic versus Meteorological Parameters ........................................................... 102
Figure 3.12: Fitted relationship for U-Test Statistic and variation in wind speed ......................... 103
Figure 3.13: Fitted relationship for U-Test Statistic and variation in Solar Flux ................................. 103
Figure 3.14: Fitted relationship for U-Test Statistic and variation in Air Temperature ................. 104
Figure 3.15: Fitted relationship for U-Test Statistic and variation in Relative Humidity ................. 104
Figure 3.16: Fitted relationship for U-Test Statistic and variation in Sky Temperature ................. 105
Figure 3.17: Fitted relationship for U-Test Statistic and variation in $\Delta T$ .......................................... 105
Figure 3.18: Fitted relationship for U-Test Statistic and time between TIC images ....................... 106
Figure 3.19: Variance of SSST against sampling interval for transects from 26/7/95 to 2/8/95 from MuBEx 95 ..................................................................................................................................................119
Figure 3.20: Variance of SSST against sampling interval for transects from 3/8/95 to 12/8/95 from MuBEx 95 ..................................................................................................................................................120
Figure 3.21: Variance of SSST against sampling interval for transects from 12/8/95 to 23/8/95 from MuBEx 95 ..................................................................................................................................................121
Figure 4.1: F value means of SSST comparisons for individual months .................................... 135
Figure 4.2: Z score means for SSST comparisons for individual months ......................138
Figure 4.3: Comparison of $F$ (top) and $Z$ (bottom) score means for SSST pairs from individual months .................................................................................................................................................140
Figure 5.1: Global climate impacts of El Niño (NOAA Climate Prediction Centre)........147
Figure 5.2: Global climate impacts of La Niña (NOAA Climate Prediction Centre) .......148
Figure 5.3: Equatorial Model SST output January 1981 to July 1991 (BSST data run) ......152
Figure 5.4: Equatorial Measured BSST January 1981 to July 1991 ...............................152
Figure 5.5: Model SST minus Measured BSST for the period January 1981 to July 1991...152
Figure 5.6: Modelled SST minus modelled climatology for 1981-95 (BSST -driven model run) .......................................................................................................................................................154
Figure 5.7: Observed BSST minus observed BSST climatology for 1981-95...............154
Figure 5.8: Modelled SST minus modelled climatology for 1981-95 (BSST -driven model run) .......................................................................................................................................................156
Figure 5.9: Modelled SST minus modelled climatology for 1981-95 (SSST -driven model run) .......................................................................................................................................................156
Figure 5.10: Difference in anomalous SST produced by BSST and SSST-driven runs for 1981-95 ........................................................................................................................................................................156
Figure 5.11: Equatorial Model SST output August 1991 to July 1995 (BSST data run) ....158
Figure 5.12: Equatorial Model SST output August 1991 to July 1995 (SSST data run) ....158
Figure 5.13: BSST minus SSST model run output ............................................................158
## LIST OF TABLES

Table 1.1: Emissivity and radiance considerations for radiometric measurement of SST .......... 17  
Table 2.1: Radiometers deployed in MuBEx ............................................................................. 42  
Table 2.2: Instrumentation deployed on the Dai-Ni Misago ................................................................................. 43  
Table 2.3: Data availability during MuBEx ‘95 ........................................................................... 46  
Table 2.4: Range of measurements recorded during MuBEx ‘95 ................................................... 46  
Table 2.5: Tasco radiometer calibration curve coefficients for different air temperatures ...... 53  
Table 2.6: Coefficients for the functions fitted to the solar flux calibrations ............................. 54  
Table 2.7: Results of noise offset determination ............................................................................ 64  
Table 3.1: Summary of calculations for one-way F-Test, unequal sample sizes; \( t_{i*} = \sum_j x_{ij} ; T = \sum_i t_{i*} = \sum_j x_{ij} \) (from Lindman, 1974) .......................................................................................... 82  
Table 3.2: Mean meteorological variations for TIC image pairs which produce F values that reject or confirm (shaded) the null hypothesis that the SSSTs in each image are from the same population (99.5% confidence level (F = 7.879)).......................................................... 85  
Table 3.3: Mean meteorological values for TIC image pairs which produce F values that reject or confirm (shaded) the null hypothesis that the SSSTs in each image are from the same population (99.5% confidence level (F = 7.879)) ......................................................................... 86  
Table 3.4: Equations fitted for F against observed variation in physical parameters............. 90  
Table 3.5: Mean meteorological variations for TIC pairs which fail & pass (shaded) the U-Test 99.5% confidence of populations being dissimilar (Z = 2.8) (150 cycle scheme)... 96  
Table 3.6: Mean meteorological variations for TIC image pairs which fail & pass (shaded) the U-Test 99.5% confidence level for dissimilar populations (Z = 2.8) (5/2 cycle processing scheme).............................................................................................................................................. 97  
Table 3.7: Mean meteorological values for TIC image pairs which fail & pass (shaded) the U-Test 99.5% confidence level for dissimilar populations (Z = 2.8) (150 cycle processing scheme).................................................................................................................................................. 98  
Table 3.8: Mean meteorological values for TIC image pairs which fail & pass (shaded) the U-Test 99.5% confidence level for dissimilar populations (Z = 2.8) (5/2 cycle processing scheme).................................................................................................................................................. 99  
Table 3.9: Equations fitted for U against observed variation in physical parameters 101  
Table 3.10: Change in F and U-Test values after application of sunglint-detection algorithms on TIC data ............................................................................................................................................. 107  
Table 3.11: Matrix of relationships derived between the variation in observed physical parameters ............................................................................................................................................. 108
Table 3.12: Categorisation of trend for standard deviation (σ) against sampling interval for 
MuBEx 95 Brightness Temperature data ................................................................. 122
Table 4.1: ATSR-1 Spectral Channels ................................................................................. 127
Table 4.2: F-Test results summary for all combinations of SSST pairs ......................... 133
Table 4.3: F-Test results summary for comparisons of SSST pairs for particular months ... 134
Table 4.4: Summary of F-Test results for direct comparisons of BSST and SSST .......... 135
Table 4.5: U-Test results summary for all combinations of SSST pairs ......................... 136
Table 4.6: U-Test results summary for comparisons of SSST pairs for particular months ... 138
Table 4.7: Summary of U-Test results for direct comparisons of BSST and SSST .......... 139
1. INTRODUCTION

1.1 Aims and Objectives

The aim of this study is to compare skin sea surface temperature (SSST) data from satellite and \textit{in situ} measurements with coincident bulk skin sea surface temperature (BSST) data. The SSST is defined in this work as being the temperature of the top 10-50\textmu m of the ocean. The BSST is defined as the temperature measured at depths from 0.1 to 10m below the surface. Detailed definitions of these terms are discussed in section 1.4. The reason for comparing these parameters is to ascertain whether any modification will be needed to climate models to allow SSST to be used as an input and to examine the implications of SSST variability for validation of SSST measured by satellite instrumentation.

This thesis focuses on two aspects of satellite and \textit{in situ} sea surface temperature measurements. The first is the dependency of the temporal and spatial variability of small-scale (less than 1km) \textit{in situ} SSST on local meteorological parameters. Two types of statistical tests are utilised to quantify SSST variability. These tests operate by analysing two test data sets to ascertain whether they are from the same population. The implications of this variability for the validation of satellite SSST measurements using \textit{in situ} SSST data is investigated.

The second aspect of this study focuses on satellite SSST measurements and co-incident large-scale (1.5 degree box) BSST data. The temporal and spatial variability of satellite SSST is investigated and compared with that of small-scale SSST using the same statistical tests applied in the study of small-scale SSST. This analysis is performed in order to investigate whether variability is the same as spatial and temporal scales are increased. If not, this has implications for climate models. There may be errors introduced if the model needs to reflect small-scale SSST variability, which is not contained within the large-scale data sets utilised by the model. This would require the model to be adjusted to take account of small-scale variability.

The variability of satellite SSST is then compared to the variability of large-scale BSST in order to study whether modifications are needed to climate models to enable them to use satellite SSST as an input. The data sets are then used to drive an ocean climate model.
The sea surface forms the boundary between the ocean and the atmosphere. The oceans cover the majority of the Earth’s surface, so the coupled behaviour of the ocean and the atmosphere is a major factor in climate study. The temperature of the sea surface plays an important role in determining the heat flux between the oceans and the atmosphere and is therefore a key component in research into climate change.

BSST measurements recorded at depths from a few centimetres to a few metres (Ma et al, 1994) have historically been used in climate models. These are measured from buoy and ship platforms and have been the mainstay of the long-term climate records. As a result, parameterisations in the climate models are tuned to the use of BSST.

Over the last few decades developments in infrared radiometry have allowed the SSST, corresponding to upper 10-50μm of the ocean surface, to be measured using satellite and in situ radiometric instrumentation. Radiometry on satellite platforms has allowed SSST to be measured on a global scale, providing far greater spatial and temporal coverage than earlier in situ data. More recently the Along Track Scanning Radiometer (ATSR) instrument on the ERS series satellites has begun to produce sea surface temperature measurements to the 0.3K accuracy for a 0.5 degree by 0.5 degree box demanded by the World Climate Research Programme in 1985. Radiometers developed for in situ measurement of SSST are also beginning to meet the 0.1K accuracy criterion needed to validate satellite instruments such as ATSR.

Satellite-measured SSST requires in situ validation to ensure that atmospheric correction and cloud detection algorithms applied to satellite data produce accurate SST. Two aspects of validation are examined: the need for more validation data and the implications of small-scale SSST variability on validation measurements.

Purpose-built radiometers for in situ deployment are expensive and therefore few in number. A relatively low-cost commercially available radiometer, the Tasco THI-500L, is evaluated as an instrument to be used to widen the availability of in situ SSST data for validation purposes. An imaging radiometer, the NEC TH1101, is used to obtain brightness temperature and SSST measurements. These data are employed to study the minimum time and area required for an in situ validation measurement to be likely to be a true reflection of the SSST in a satellite pixel.
1.2 Infrared Radiometry

The detection of radiation emitted by a body in the infrared region of the electromagnetic spectrum is used as a non-contact method of measuring the temperature of that body. The infrared region is used principally because bodies at temperatures of 270-300K (typical of terrestrial sea and land surfaces) emit most energy at these wavelengths. This is shown by the Planck emission spectra for a black body (those of unity emissivity) at various temperatures displayed in Figure 1.1. A black body is a body that is a perfect emitter over all wavelengths. The emission spectra from a body at $T$ deg. Kelvin is described by Planck’s radiation law:

$$B_{\lambda T} = \frac{C_1}{\lambda^5 \left[ \exp \left( \frac{C_2}{\lambda T} \right) - 1 \right]}$$

Equation 1.1

Where $B_{\lambda T}$ is the radiant flux density per unit bandwidth centred on wavelength $\lambda$, leaving unit area surface of black body at temperature $T$. $C_1$ and $C_2$ are constants with values:

$C_1 = 3.74 \times 10^{-16}$ Wm$^2$

$C_2 = 1.44 \times 10^{-2}$ m deg K

$B_{\lambda T}$ has units Wm$^{-2}$m$^{-1}$. Wavelengths are often expressed in $\mu$m so to obtain $B_{\lambda T}$ in Wm$^{-2}$m$^{-1}$, Equation 1.1 must be multiplied by $10^6$.

![Figure 1.1: The Planck emission spectra for a black body at various temperatures](image)

Emissivity is defined as the ratio of $B_{\lambda T}$ for a real surface at a given temperature to $B_{\lambda T}$ for a black body at the same temperature. The sea surface is not a black body. It has an emissivity that is dependent on wavelength, look angle and weakly on temperature. This requires that
Equation 1.1 is multiplied by the emissivity, $\varepsilon_\lambda$, at the radiometer look angle. Non-unity emissivity means that a fraction of the downwelling radiation equivalent to $1 - \varepsilon_\lambda$ is reflected by the sea surface. This means that the signal detected by the radiometer contains a reflected sky component as well as the radiance emitted from the sea surface (Figure 1.2).

![Figure 1.2: Composition of the radiometric signal from the sea surface](image)

If the reflected component of the measured radiance is large in comparison to the emitted component then there is a greater likelihood of error in obtaining the emitted radiance. Therefore the range of the electromagnetic spectrum that the radiometer is sensitive to must be such that the reflected radiance is minimised. This is achieved by comparing the $\varepsilon_\lambda$ values with the downwelling radiance from the sky. Figure 1.1 shows the Planck spectra for black bodies at various temperatures. Absorption of this radiation by the atmosphere occurs at various wavelengths (Figure 1.3) by, for example, aerosols, ozone and oxygen. However there are several wavelength "windows" (for example 3.5μm, 10.5μm and 12.0μm) where atmospheric transmission is high. The spectra for typical sea surface temperatures ($T = 300K$) and sky temperatures ($T = 250K$) mean that the radiances are similar over these windows. However, Table 1.1 shows how $\varepsilon_\lambda$ of the sea surface is high over these wavelengths. This implies that only around 2% of the signal is reflected sky radiance, thereby minimising the errors in SSST retrieval.

Table 1.1 also shows how the downwelling solar radiance and upwelling radiance vary with wavelength. The values of $\varepsilon_\lambda$ were obtained by experiment (Masuda et al., 1988). At the mid-infrared region the reflected solar radiation component dominates the signal from the sea surface. In the thermal infrared region the reflected radiance is of a similar magnitude to that emitted from the sea surface. This has implications primarily for satellite SST retrieval. The signal in the 3.5μm band can have a strong component from the specular reflection of solar radiation and therefore data from mid-infrared bands are not generally used for retrieval of SSST during the day.
Table 1.1: Emissivity and radiance considerations for radiometric measurement of SST

<table>
<thead>
<tr>
<th></th>
<th>Mid Infrared $(\lambda = 3.5\mu m)$</th>
<th>Thermal Infrared $(\lambda = 11.0\mu m)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Emissivity of sea surface</td>
<td>0.971</td>
<td>0.993</td>
</tr>
<tr>
<td>(from Masuda et al., 1988)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Downwelling solar radiance</td>
<td>$=10^5$ W m$^{-2}$ $\mu m^{-1}$</td>
<td>$=10^4$ W m$^{-2}$ $\mu m^{-1}$</td>
</tr>
<tr>
<td>(from body at $T = 6000K$)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reflected component of</td>
<td>$=3 \times 10^4$ W m$^{-2}$ $\mu m^{-1}$</td>
<td>$=10^2$ W m$^{-2}$ $\mu m^{-1}$</td>
</tr>
<tr>
<td>signal from sea surface</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Radiance emitted from sea</td>
<td>less than 1 W m$^{-2}$ $\mu m^{-1}$</td>
<td>$5 \times 10^1$ W m$^{-2}$ $\mu m^{-1}$</td>
</tr>
<tr>
<td>surface at $T = 300K$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Radiometric measurements of SST are thus concentrated in the thermal infrared region of the electromagnetic spectrum. Satellite radiometry must also take into account absorption by the atmosphere as shown in Figure 1.3. The two “windows” at 10.5μm and 12.0 μm in the thermal infrared which have little absorption are widely used by satellite infrared radiometers to measure radiance from the sea surface.

![Atmospheric transmission in the visible and infrared](image)

Figure 1.3: Atmospheric transmission in the visible and infrared

Radiance values from these radiometers still require correction for atmospheric absorption. Levels of aerosols and especially water vapour fluctuate particularly with latitude. Figure 1.3 shows how atmospheric transmission varies with wavelength and latitude, where transmission is defined as the ratio of energy transmitted by a body to that incident on it. The Advanced Very High Resolution Radiometer (AVHRR) on the NOAA series satellites uses algorithms based on using both thermal infrared “windows” to get a “dual band” look at the same target in order to correct for the spectral dependence of atmospheric absorption (Barton, 1985). The Along Track Scanning Radiometer (ATSR) on the ERS series satellites uses both the dual
band system plus a dual look system, which scans the same part of the sea surface from two different angles. This allows the radiance through two different path lengths to be measured. Radiative transfer models can then be developed (Zavody et al., 1995) to correct for transmission through the atmosphere.

Radiometers deployed to get *in situ* measurements of SSST have been generally of the same narrow band type as their satellite counterparts. (Nightingale 1997; Barton et al. 1995). However, the short (= 5-10m) atmospheric path length between a ship based radiometer and the ocean surface means that absorption will be greatly reduced. Therefore, a broad band radiometer can be used for *in situ* observations providing that the band does not extend to shorter wavelengths where the reflected component will dominate. Instruments sensitive to radiation in the 8-13μm wavelength region fulfil these criteria.

Detailed descriptions of the processes for SSST retrieval from the *in situ* and satellite radiometry used in this study are given in Chapters 2 and 4 respectively.

1.3 Considerations for validation of satellite SST measurements using *in situ* radiometry

Satellite SSST has superior spatial (global) and temporal (≈1.5 days) coverage to that of *in situ* “point” measurements from ships and buoys. The coverage is far more consistent than *in situ* data as the satellite orbit will usually be configured so that it passes over a particular location at regular intervals. It also provides the data at a lower cost than using *in situ* methods to achieve the same coverage.

The use of satellite platforms for measuring SSST has required the development of SSST retrieval algorithms to derive, and attempt to correct for, the effect of the atmosphere. Earlier data from satellite instruments such as the Advanced Very High Resolution Radiometer (AVHRR) on the NOAA series satellites used Bulk SST from ships and buoys to validate and develop retrieval algorithms (Llewellyn-Jones et al. 1984, Brown et al. 1985, Yokoyama and Tanba 1991). These showed that monthly mean SST accurate to 0.5K could be derived from AVHRR data (Folland et al., 1993).

In 1985 the World Climate Research Programme defined a 0.3K SSST accuracy for a 0.5 degree by 0.5 degree area of ocean surface. In order to meet this requirement, the Along Track Scanning Radiometer (ATSR) series of instruments were designed with the principal
purpose of producing accurate long-term measurement of SSST. Two of these instruments have been launched: ATSR/1 in 1991 on the European Space Agency ERS-1 satellite and ATSR/2 in 1995 on ERS-2. The third instrument in this series, the Advanced Along track Scanning Radiometer (AATSR) features further enhancements and is due for launch in 2000. The instruments make use of an improved design using on board black bodies for in-orbit internal calibration. The SSST is retrieved by application of an algorithm using coefficients derived from a radiative transfer model (Zavody et al., 1995). These features produce SSST accurate to 0.3K or better over the 0.5 degree latitude by 0.5 degree longitude spatial resolution of an Averaged SST product (ASST) box (Mutlow, 1994) when compared to in situ SSST measurements. Aerosol contamination of the atmosphere from the Mount Pinatubo eruption in 1991 has resulted in work to refine this algorithm for early ATSR data. Results for this algorithm for ATSR data around the British Isles show SSST to a precision of 0.08K can be achieved (Brown et al., 1997).

The radiometric instrumentation used to get in situ data needs to be of a similar accuracy as the satellite based counterparts for which validation data is being obtained, otherwise measurement errors will not allow the satellite SSST to be validated to the design accuracy of the satellite radiometry. Ship-mounted narrow-band radiometers are normally purpose-built for the task of obtaining measurements to validate satellite SSST. Examples of these include the Scanning Infrared Sea Surface Temperature Radiometer (SISTeR) (Nightingale, 1997) and those built by the CSIRO laboratory in Australia (Barton et al., 1995). The purpose-built nature of these radiometers limits their deployment due to the cost per unit (typically £20,000 to £60,000) and expertise required to operate them. However there is an extensive commercial market in broad-band radiometry for industrial applications such as product regulation and environmental temperature monitoring. The performance of these instruments continues to be refined and improved to meet the needs of the commercial user.

An example of such a radiometer, the Tasco THI-500L, was deployed during the 1995 Mutsu Bay Experiment (MuBEx), off the northern end of Honshu Island, Japan (see section 2.4 in Chapter 2). The accuracy and performance of the Tasco THI-500L in measuring SSST was assessed as part of this experiment. Laboratory calibration tests of this radiometer were also undertaken. The THI-500L was chosen with a view to portability, ease of deployment, low cost and accuracy for validation of satellite radiometric SSST.

Another recent innovation in radiometry is thermal infrared imaging “cameras”. These can produce an image of the sea surface allowing the investigation of small-scale temporal and
spatial variability. The commercially available NEC TH1101 thermal infrared camera (TIC) was deployed during the 1995 MuBEx. This has a similar broad band filter profile to the Tasco THI-500L (see Figure 2.2 and Figure 2.3) and produces a 207 × 255 pixel image with each pixel representing approximately 1cm² of sea surface area when the camera is mounted 5m above the sea surface. The accuracy and performance of the NEC TH1101 for in-shore deployment during the 1995 MuBEx was assessed by the Japanese MuBEx group and their results are referred to in sub-section 2.4.4.2 of Chapter 2.

The TIC is a particularly useful tool for validation research in the context of establishing recommended criteria for the time and coverage needed for in situ validation measurements. Previous ATSR validations using in situ radiometers (Mutlow et al., 1994; Barton et al., 1995) have used a one minute average of the radiance at the time of the satellite overpass as the validation measurement. This validation measurement is normally extracted from a set of radiometer data for a transect across part of the ATSR field of view. Obtaining data before (and sometimes after) the ATSR overpass allows the radiometer and recording equipment to be checked.

The use of a one minute “snap shot” of the in situ radiometric SSST as a validation point raises the issue that the area of the ocean skin observed is much less than the 1km² pixel size of the ATSR image. Typically for a ship travelling at 1ms⁻¹, operating a radiometer which has a spot size on the ocean surface of width 0.5m, a one minute integration time will yield the radiance over 60m × 0.5m = 30m². This represents 0.003% of the nominal ATSR nadir pixel area. The ATSR has an integration time of 75μs for each pixel. This means that (75 × 10⁻⁶/60) × 100 = 0.0001% of the in situ data was acquired at the same time as the ATSR data. This introduces errors into the in situ SSST measurement if the area of ocean surface being observed is atypical of the mean SSST across the satellite radiometer pixel.

In Chapter 3 the TIC is used to investigate the spatial and temporal variability of SSST for a field of view similar to that of a typical in situ validation radiometer.
1.4 The Skin-Bulk Temperature deviation

1.4.1 Definition of the skin-bulk temperature deviation

The development of infrared radiometers has allowed measurement of the “skin” of the sea surface. The “skin” can be defined as the top 10-50μm of the ocean, this being the optical depth of sea water in the thermal infrared (wavelengths of 10-12μm) (Dobbs, 1985)

The bulk sea surface temperature is that measured by contact thermometry below the skin. There is no one defined depth for the BSST, observations generally being recorded at depths from 0.1 to 10m below the surface. It is important to note that the depth at which the BSST data is measured should be known for the values of the skin bulk temperature deviation. The skin-bulk temperature deviation is defined as:

$$\Delta T = T_{\text{skin}} - T_{\text{bulk}}$$

Equation 1.2

This definition is the one agreed at the Combined Action to Study the Ocean's Thermal Skin (CASOTS) I workshop of 1996 (Donlon et al., 1998) in Southampton, UK. Some authors prefer to use $T_{\text{bulk}} - T_{\text{skin}}$ as the definition for $\Delta T$ but this work will use the above CASOTS definition.

The difference between $T_{\text{bulk}}$ and $T_{\text{skin}}$ is governed by a variety of physical processes at the ocean-atmosphere interface. The understanding of the effect of these processes is therefore essential to the comprehension of the behaviour of the bulk-skin temperature difference $\Delta T$ and the possible impacts of using SSST, as opposed to, or to complement BSST in climate models.

1.4.2 Why is there a “cool skin”?

The existence of a “cool skin” whereby the BSST is greater than the SSST, was first postulated by Bruck (1940) and Woodcock (1941). This leads to negative values of $\Delta T$ as given in Equation 1.2. Radiometric measurements of SSST (Saunders, 1967; Katsaros, 1977; Grassl 1976; Schluesel et al., 1990) showed a differential of up to 1 degree Kelvin between the skin and simultaneous BSST. The BSST in these examples was obtained via thermometric measurements of sea temperature from a bucket dipped into the sea over the side of a ship or by contact thermometry.
There are several processes by which energy is transported in the uppermost layers of the ocean. The net heat flux, $Q$, across the atmosphere/ocean interface is given by:

$$Q = E + H + S + L$$

**Equation 1.3**

This is by convention negative for a heat loss from the ocean. $E$, (Latent heat) and direct thermal transfer, $H$, (Sensible heat) are described below. The net incoming short-wave solar radiation is $S$. The net long wave radiative heat loss, $L$, is given by

$$L = L_{SKY} - L_{SUR} = e_{SKY}^4 T_{SKY}^4 - e_{SUR}^4 T_{SUR}^4$$

**Equation 1.4**

Where $T_{SKY}$ is the sky temperature; $T_{SUR}$ is the sea surface temperature; $e_{SKY}$ and $e_{SUR}$ are the spectrally integrated emissivity of the sky and sea surface respectively. Since the sky temperature is cooler than the sea temperature, $L$ is almost always negative. Non radiative heat loss occurs through evaporative cooling, $E$, (Latent heat) and direct thermal transfer, $H$ (Sensible heat). These turbulent fluxes can be calculated by parameterising drag, heat transfer and evaporation coefficients in terms of local meteorological factors. These are described in the following formulae (Smith, 1988):

Latent heat flux

$$E = C_L \rho L u (q_a - q_s)$$

**Equation 1.5**

Sensible heat flux

$$H = C_H \rho c_p u (\theta - T_{SUR})$$

**Equation 1.6**

Wind Stress

$$\tau = C_D \rho u^2$$

**Equation 1.7**

Where $C_L$ and $C_H$ are the evaporation and heat flux coefficients; $C_D$ is the drag coefficient; $\rho$ is the air density; $u$ the mean wind velocity; $q_a$ and $q_s$ are the water vapour mixing ratios of the sea surface and atmosphere; $c_p$ is the specific heat of air at a constant pressure; $L$ the latent heat of evaporation; $T_{SUR}$ is the sea surface temperature; $\theta$ is the overlying potential air
temperature. \( \theta, q_a \text{ and } u \) are normalised to a reference height of 10m. Values of the coefficients are tabulated by Smith (1988). These are consistent with empirical data, which depend on wind speed, potential air temperature and sea surface temperature. For example, it can be seen from Equation 1.5 and Equation 1.6 that the Latent and Sensible Heat fluxes will both increase with wind speed.

The exchange of flux at the atmosphere/ocean interface is restricted to molecular conduction processes unless the skin is destroyed by breaking waves or spray. This is because turbulent mixing near the surface is limited by the surface tension of the ocean skin (Hasse and Liss, 1980). Above and below this skin, turbulent eddy mechanisms transport heat.

The heat exchange processes given in Equation 1.3 determine temperature gradient and depth of the ocean skin layer. Water has a high emissivity in the thermal infrared (Masuda et al., 1988). At a typical range of SSST temperatures of 273K – 310K, the peak of the emission spectrum is in the thermal infrared (Figure 1.1). These two factors mean that longwave flux is radiated from, and absorbed in, the top 10-20\(\mu\)m of the ocean, thereby cooling or warming the skin layer. The skin temperature therefore responds most rapidly to the generally upward (cooling) net longwave radiance.

There is also generally a net transfer of latent and sensible heat from the ocean to the atmosphere with cooling of the surface by latent heat (of evaporation) often dominating the cooling by sensible heat and longwave radiance. Thus the top millimetre of the surface where conduction dominates as the form of heat transfer within the water, will support a temperature gradient due to the low conduction coefficient of water and the lack of turbulent processes. The various processes leading to the cool skin are illustrated in Figure 1.4.
1.4.3 Non-"cool skin" conditions

The effect of the short wave radiation component in Equation 1.3 has not been considered yet in this discussion. Schmidt (1908) calculated the percentage of solar radiation absorbed in clear water at different path lengths (Figure 1.5). These percentages can be greater if we consider sea water containing suspended matter (Gemmrich and Hasse, 1992). In the top millimetre of the ocean, this warming is not normally greater than the cooling from the latent, sensible and longwave terms. However under low wind speed conditions, turbulent mixing of the upper 0.5-1.0m of the ocean is less active, and the cooling of the skin by Latent and Sensible heat fluxes is reduced, leading to the formation of a diurnal thermocline. This "warm skin" phenomenon was observed by Yokoyama et al. (1995) (Figure 1.6) using a thermistor chain attached to a buoy. The 4.0K magnitude of this skin minus bulk temperature difference may be erroneously high due to solar heating of the near surface thermistors, since all temperature measurements were acquired using contact thermistors. However, the data clearly shows the breakdown of the thermocline once wind speed exceeds 2ms⁻¹. This suggests that warm skin phenomena are localised to calm conditions and high solar insolation.
Figure 1.5: Absorption of solar radiation in clear water (data from Schmidt, 1908)

Figure 1.6: Formation and destruction of a diurnal thermocline (Yokoyama et al., 1995)

The presence of solar heating during the day leads to more positive values of $\Delta T$. The mean value of $\Delta T$ was found by Schluessel et al. (1990) to be -0.11K during the day and -0.3K at night. This is because during the night, the sea is usually warmer than the air, leading to strong cooling of the skin in the thermal infrared. During the day cooling from turbulent and thermal infrared fluxes are offset by warming by solar flux. The dependence of the turbulent fluxes on wind speed (Equation 1.4 and Equation 1.5) results in calm conditions producing less cooling of the skin and less turbulent transport of heat from the upper layers of the ocean.
to deeper water. This allows more of the heating from short wave solar flux to be retained near the surface producing the diurnal thermocline. Higher wind speed conditions during the day increase the cooling from turbulent fluxes and mixing of the upper layers of the ocean, resulting in a zero or negative value of $\Delta T$, depending on the magnitude of the wind speed and levels of solar flux.

Observations of $\Delta T$ in high wind speed conditions (Donlon and Robinson, 1997) suggest that the skin-bulk temperature deviation becomes negligible at wind speeds of $>10\text{ms}^{-1}$. This would be consistent with the continuous destruction of the skin layer by breaking waves, foam and spray. Studies have shown that the skin layer needs 10 to 12s to re-form after being destroyed by such phenomena (Ewing and McAlister, 1960; Clauss et al., 1970). This may not occur under continuously high wind speeds.

The presence of surface films or slicks on the ocean surface will also affect the magnitude of $\Delta T$ by influencing the value of SSST (Kroptokin et al., 1978; Katsaros 1980). These can be man made or naturally occurring features. It cannot be readily predicted whether slicks will increase or decrease the measured SSST. They can reduce the emissivity of the sea surface, which would result in radiometric measurements of SSST being cooler than the true SSST if the higher “normal” sea water emissivities were used in the SSST retrieval. The slick will reduce turbulence in the upper layer induced by wind stress, leading to a thickening of the skin. The slick itself provides an extra insulating layer through which only molecular conduction facilitates heat exchange. These two effects will both lead to a cooler SSST if the net heat flux is from the sea to the atmosphere.

Gas exchange and evaporation can be inhibited by the presence of slicks. This leads to a warmer SSST by reducing the heat flux transfer to the atmosphere via the Latent Heat term. However Katsaros (1977) argues that organic slicks usually do not inhibit evaporation and will give rise to a cooler skin due to the slick reducing turbulence in the upper layers thereby thickening the conductive layer. Johnston (1997) finds that two large man-made oil slicks (from tanker spills) decrease the SSST observed from satellite platforms by 0.4 - 1.0K. Small-scale slicks observed in inshore waters in Mutsu Bay, Honshu, Japan using in situ radiometry were found mostly to have no effect on SSST and it was speculated that this is because the different effects these slicks have on the SSST cancel out.
1.4.4 Parameterisations and magnitude of the skin-bulk temperature deviation

Several campaigns have been undertaken to acquire \textit{in situ} skin-bulk temperature data sets. Recorded values for the skin-bulk temperature deviation lie in the range $-1.0 \text{K} \leq \Delta T \leq 4.0 \text{K}$. These data sets have been used to formulate and verify models that can predict the value of $\Delta T$. The models fall into four categories.

The first type is typified by the Saunders (1967) model which has been tested by several authors (Grassl, 1976; Schluessel \textit{et al.}, 1990; Donlon, 1994; Donlon and Robinson, 1997). This assumes that the skin of the ocean is a stagnant film through which heat is transported by molecular conduction. Above and below this layer, turbulent processes transfer heat from and to the stagnant layer with an infinite turbulent transport coefficient. Assuming that most of the skin-bulk temperature gradient is across the stagnant (viscous) layer and by obtaining a thickness for this layer by dimensional analysis, Saunders postulated:

$$\Delta T = \frac{-\lambda (v / \kappa) Q}{u^* c_p \rho}$$

\textbf{Equation 1.8}

Where $v$ is the kinematic viscosity; $\kappa$ is the coefficient of molecular diffusion; $Q$ is the net longwave, sensible and latent heat flow measured across the stagnant layer; $u^*$ is the friction velocity of the upper ocean; $c_p$ and $\rho$ are the specific heat capacity and density of sea water respectively; $\lambda$ is a wind dependent dimensionless coefficient. The model has been modified to include a solar absorption term, which allows it to predict $\Delta T$ for daytime conditions (Donlon and Robinson, 1997). An assumption is made that the solar warming only contributes to the radiometric signal in the thermal sublayer. This is smaller than the stagnant (viscous) layer and the thickness of the thermal sublayer is given by:

$$\delta = K / u^*$$

\textbf{Equation 1.9}

Where $K$ is the thermal conductivity. Saunders (1967) proposed that values of $\lambda$ should be in the range 5 to 10. Although some authors have found agreement with this (Paulson and Simpson, 1981) others have empirically derived values for $\lambda$ to fit their data (Grassl, 1976; Schluessel \textit{et al.}, 1990; Donlon, 1994; Kent \textit{et al.}, 1996; Fairall \textit{et al.}, 1996). Most of these $\lambda$
values are wind dependent although Kent et al., (1996) propose a fixed value of $\lambda=7.8$, and other authors (Fairall et al., 1996; Wu, 1995) suggest that $\lambda$ is invariant at high wind speeds.

The Hasse (1971) scheme is an example of the second type of model. This treats the surface as a rigid boundary and assumes that the efficiency of turbulent diffusion of heat is a function of distance from the boundary. The temperature gradient can then be retrieved by using a known heat flux and a function of the turbulent diffusion coefficient of heat. By plotting $\Delta T$ against $Q^*$, the ratio of heat flux through the surface (excepting solar flux), to $U$, the mean wind speed at 4m above the sea surface, Hasse found:

$$\Delta T \propto \frac{Q^*}{U}$$

Equation 1.10

Hasse then included a term to account for the solar flux, $Q_s$, to derive the following relationship:

$$\Delta T = C_1 \frac{Q^*}{U} + C_2 \frac{Q_s}{U}$$

Equation 1.11

Where $C_1$ and $C_2$ are derived coefficients. $C_1$ varies slightly with $z$, the depth of the bulk measurement while $C_2$ varies more strongly with $z$ due to the logarithmic absorption profile of solar flux by sea water.

The third type of model is based on surface renewal. The surface is thought of as discrete elements, which will cool and become denser, thereby being replaced by warmer elements from beneath after a finite time. Thus heat is exchanged between the surface and deep layers of the ocean. The intensity of the turbulence governs the amount of time an element will remain at the surface. There are several examples of this type of model (Eifler, 1992; Soloviev and Schluessel, 1994; Wick, 1995).

The fourth type of model is the use of a statistical fit to the in situ data to derive a parameterisation for $\Delta T$. A stepwise multiple regression was used by Schluessel et al. (1990) to obtain two parameterisations, one for day time and one for night time.
Donlon and Robinson (1997) and Kent et al. (1996) tested many of these models against *in situ* data. Donlon and Robinson (1997) use measurements made in the Atlantic Ocean between latitudes 52°N and 20°S. All the above models were found to be unable to account for the range of conditions encountered during the measurements (e.g. high wind speeds, several different climate regions). The regional nature of the observations from which the models were derived was given as a possible cause of the models being unable to predict $\Delta T$ over a range of climatologies. The exception to this was the parameterisation of Wick (1995) which still produced high scatter when correlated against observations of $\Delta T$ but was able to incorporate the large variability of these observations.

Figure 1.7 shows the results of Kent et al. (1996) using observations of $\Delta T$ made in the Atlantic Ocean between latitudes 34°N and 46°N to test the models. Cases where a diurnal thermocline could be formed were excluded by ignoring data where the solar flux exceeded the empirically determined value of 800 Wm$^{-2}$. The Saunders model was applied without using the solar flux term in determining $Q$ and found to agree well with *in situ* measurements of $\Delta T$ at wind speeds of 3-7 ms$^{-1}$. 
1.4.5 The dependency of the variability of SSST on local meteorological parameters

In Chapter 3 the dependency of the variability of SSST on parameters that determine the net ocean-atmosphere heat flux, such as solar flux, wind speed, air temperature, humidity and longwave downwelling flux, is examined. This is to ascertain whether the variability of SSST on small scales is driven by the same factors that control the net ocean-atmosphere heat flux. Data from the Thermal Infrared Camera deployed during the 1995 MuBEx is used for this analysis. Two statistical tests for similarity of test populations, the analysis of variance F-Test and Mann-Whitney U-Test, are used as measures of SSST variability. The behaviour of
SSST variability is further explored in Chapter 4. Here, the statistical tests are applied to large-scale SSST measurements of the equatorial Pacific Ocean from the ATSR-1 instrument. This is in order to investigate whether the variability of SSST has similar properties on small (metre) and global scales.

1.5 Climate models and the use of satellite-derived SST

Traditionally BSST measurements have been used in climate modelling studies (Barnett et al., 1994; Ma et al., 1994; Seager and Blunenthal, 1994; Dewitte and Perigaud, 1996), with BSST providing the initial conditions for a climate model. Alternatively, BSST has been used to “force” the model or as a check to compare the model’s predicted SST with observed SSTs. Satellite and in situ radiometric SSST data have only become available in the last few decades. Climate models are therefore mostly tuned to use BSST as this data has been available since the last century, being measured by ships or buoys using contact thermometry.

The use of satellite derived SSST data maybe more beneficial than BSST since the SSST is a key factor in determining the heat flux across the ocean-atmosphere boundary. Satellite SSST also provides more frequent and extensive global coverage than the point measurements of ships and buoys used to measure BSST. Coverage of BSST data can be particularly poor in areas of ocean away from main shipping lanes or dedicated buoy networks. Conversely it must be noted that satellite SSST is only available in the absence of cloud and therefore the satellite SSST data set is biased towards measurements obtained during clear sky conditions. Thus, it may be that satellite SSST should be used to complement, rather than replace, existing BSST data for use in climate modelling. The use of in situ SSST data may allow a better understanding of how the SSST behaves during cloudy conditions, enabling the “gaps” in satellite SSST data caused by cloud contamination to be filled in and the bias of the data towards clear-sky conditions to be accounted for.

The Reynolds (1988 and 1993) SST is often used as part of climate modelling studies (e.g. Dewitte and Perigaud, 1996). The Reynolds analysis uses a combination of in situ bulk and satellite-derived SST to produce a global SST data set. However this analysis uses AVHRR data derived using retrieval algorithms tuned to BSST validation measurements. So although the Reynolds SST analysis contains an element of satellite radiometric data, it is a BSST data set. Furthermore, AVHRR is in fact detecting the radiative flux from the ocean skin. This means that an algorithm tuned to provide BSST measurements from AVHRR data will
introduce an error of the order of the uncertainty in the skin effect into the retrieved BSST value.

Another approach is to use satellite radiometric measurements as part of a scheme to retrieve the net surface flux. However use of these fluxes in climate models has been limited due to the lack of information about the thermal content in the low-level atmosphere (Eymard and Taconet, 1995).

The rarity of \textit{in situ} SSST data and the relative novelty of accurate satellite SSST data have meant that the potential of using these data in modelling studies has still to be realised. In Chapter 4 the statistical tests for variability applied in Chapter 3 are used to test whether global-scale BSST and ATSR-measured SSST data sets are statistically similar. These results are used to address the question as to whether the use of SSST data in climate models instead of, or to complement, the existing BSST data, will require any adjustment of the models.

In Chapter 5, the ATSR average sea surface temperature (ASST) product is used to drive an ocean model of the Tropical Pacific Ocean. The results are compared with model runs using the BSST data from the combined ocean atmosphere data set (COADS) to drive the model.

### 1.6 Summary

This study will explore the relationship between the variability of skin and bulk sea surface temperature. This area of research has several areas of application. Acquisition of accurate and representative \textit{in situ} SSST enables validation of corresponding satellite measured data. In Chapter 2 the accuracy of a low cost radiometer is examined with a view to utilising these instruments to widen the availability of \textit{in situ} SSST for validation purposes.

An \textit{in situ} imaging radiometer is deployed to allow investigation of the spatial and temporal variability of SSST on small-scales. This analysis is presented in Chapter 3 and has two main aims: to obtain a spatial and temporal specification for \textit{in situ} SSST measurements that is likely to yield a "typical" local SSST value for validating satellite measurements; to determine the dependency of the variability of \textit{in situ} SSST with respect to local physical parameters. The results from the latter investigation can then be compared with the dependency of $\Delta T$ on the same physical parameters to see if SSST variability and $\Delta T$ are governed by the same factors.
The analysis of small-scale SSST in Chapter 3 is then applied in Chapter 4 to large-scale SSST data from the ATSR radiometer on the ERS-1 satellite. The first aim of Chapter 4 is to compare the results of the variability of small-scale SSST from Chapter 3 to large-scale SSST variability. This is to examine whether SSST exhibits a higher variability over small temporal and spatial scales than over large-scales. The second aim of Chapter 4 is to investigate the variability of large-scale SSST with respect to that of large-scale BSST. If the two data sets exhibit different statistical variability then a resulting hypothesis can be drawn that the substitution of SSST for BSST to drive a climate model will produce a significant deviation in the model output. The hypothesis formed by the analysis in Chapter 4 is tested in Chapter 5 by forcing an ocean model with BSST and then SSST in two separate runs. The model performance for these two runs is assessed to allow a conclusion to be formed as to whether the use of SSST as a substitute for, or to complement BSST will require modification to climate models. This has important implications for validating and enhancing climate models using all available data. That is, both SSST and BSST.
2. DATA ACQUISITION AND INSTRUMENT STABILITY

2.1 The need for in situ measurements of sea surface temperature

Accurate in situ sea surface temperature (SST) measurements have several important applications. These include validating satellite SST instruments, studying small-scale ocean-atmosphere processes and providing a direct input into climate models. Validation measurements are required to be of the same accuracy as the satellite radiometers that are being validated. The current state of the art satellite radiometers such as the ATSR series of instruments are designed to have a 0.3K accuracy in measuring SST over a 0.5 degree latitude by 0.5 degrees longitude box.

Bulk SST (BSST) data is often acquired by measuring the temperature of ship engine intakes. This can cause problems with warming from the engine but the thermometry used is widely available and many ships are equipped to measure BSST. However BSST is not a true reflection of the radiometric SST measured by satellite instrumentation due to a combination of the skin effect and diurnal effect (Chapter 1). Therefore validation of satellite-measured SST using BSST will introduce errors of the order of uncertainty in the skin effect and diurnal effect.

Skin SST (SSST) data sets are less common than BSST because the radiometry required to retrieve SSST are fewer in number. These in situ radiometers are required to be both robust in oceanic conditions and to be able retrieve accurate SSST. To date, this has resulted in radiometers being purpose-built and usually requiring specialist operators, adding cost and expertise limitations on their deployment. A stable, accurate, low cost in situ radiometer that is relatively easy to operate and maintain could greatly increase the availability of validation measurements and thus the accuracy of SSST retrieval algorithms for radiometric satellite data.

2.2 Radiometers: components and terminology

The heart of a radiometer is a detector sensitive to incident radiative energy at a particular range of wavelengths (normally in the infrared) on its surface. However the detector quickly becomes saturated under continual exposure to photons and a chopper (often a rotating piece of black coloured metal) is periodically introduced into the optical path in order to obscure the detector from the incident flux and allow a zeroing of the detector. To restrict the range of
wavelengths measured still further, a filter with a known transmission spectrum is also introduced between the detector and the incident flux. A calibration target at a known temperature and with a known emission spectrum may be moved into field of view of the detector from time to time, either mechanically or by diverting the view of the detector using a mirror. Calibration targets are used to check the accuracy of the detector/filter system and are discussed further in section 2.3.

Apart from accuracy, a radiometer can also be defined in terms of several other physical characteristics. The full width at half maximum (FWHM) of the filter is a measure of the spectral response of the filter. It is given by the difference between the two extreme values of the wavelength at which the transmission of the filter is equal to half of its maximum value. The term bandpass is often synonymous with FWHM, although bandpass can be used to describe the difference between the two wavelength values for which the spectral response of the filter is equal to a particular percentage of transmission.

The sensitivity of the detector is the ratio of the response in terms of counts for a given change in blackbody target temperature. The greater the change in counts for a 1°C change in equivalent blackbody temperature, the greater the sensitivity of the channel. Sensitivity is measured in counts °C⁻¹.

2.3 Considerations for accurate in situ radiometric SSST measurements

Radiometers must be calibrated periodically to ensure that their response in terms of gain, offset and linearity is known for the retrieval of SSST. This is achieved by measuring the radiance from one or more calibration targets. The range of environmental conditions experienced at sea, plus the time that an instrument may be away from the laboratory often requires that calibration systems are deployed with the radiometer if accurate SSST is to be determined.

The calibration target should ideally be kept at a temperature similar to that of the SSST being measured so that the calibration process obtains the response of the radiometer for that particular target radiance. If two calibration sources are used then one should be kept at a temperature above that of the observed SSST range while the other is at a temperature below the SSST range. A linear response is assumed, allowing interpolation of the radiometric response for intermediate radiances. However it is often difficult to cool calibration targets to
a steady temperature below that of the ambient air temperature, especially in the field on board ship. "Cool" black body targets at temperatures below ambient air temperature show a tendency to produce condensation from water vapour in the air. This can reduce the emissivity of the black body and damage the rest of the radiometer assembly.

The cooler calibration target is therefore frequently kept at ambient air temperature and, assuming a linear response, the characteristics of the radiometer for lower radiances can be extrapolated. Changes in the response of the radiometer can be caused by factors including changes in air temperature, contamination (e.g. by salt air) of the optics, or solar heating affecting the performance of the detector and electronic components. The requirement for a 0.3K accuracy criterion means that high quality components producing precise radiance measurements must be used in the construction of the radiometer. These radiometers are generally prototype in nature and have added complexities such as scan mechanisms for switching between "external" radiance and black body targets. Purpose-built skin SST radiometers are therefore far more expensive and tend to require more maintenance than the thermometric devices used to measure BSST. These factors coupled with the fact that "contact" thermometric technology has been available far longer than "non-contact" radiometry means that in situ SSST data are scarce relative to BSST data.

Calibration sources on-board ship can be categorised into either internal (defined as either being within the radiometer assembly or located before a radiometer's external protective window if the radiometer has one) or external targets. Internal targets are mostly black bodies. These are thermally homogenous objects with near unity emissivity so that the observed brightness temperature is equal to the ambient temperature of that object. Black bodies can be moved into the field of view via an appropriate mechanism. An alternative arrangement for an internal black body is to utilise a mirror to direct the radiation from the black body onto the detector. As the internal calibration target is located within the radiometer assembly it is often well protected from contamination by airborne sea salt or other contaminants during long-term operation. Black bodies present engineering challenges to ensure they are thermally homogenous and stable to ensure that the thermometric data from within the black body accurately reflects the true black body temperature. They must also have a known, uniform emissivity and accurate thermometry.

External calibration targets can be either a "stirred bucket" of seawater or a black body. The stirred bucket is a tank of water that either stirred via an appropriate mechanism, or more
often it is continually replenished by sea water pumped directly from the ocean. These measures are designed to break the skin of the water in the bucket so that the radiometer is detecting the radiance from the bulk water temperature (measured by a contact thermometry) and not a skin temperature. The stirred bucket is periodically moved into the field of view of the radiometer to allow the calibration measurement to be recorded. The stirred bucket can be cumbersome and slow to deploy as a calibration target because they must be moved in and out of the view of the radiometer along a rail or by hand. If external black body calibration targets are used they are more susceptible to contamination during long deployments at sea.

External calibration sources have the advantage of calibrating the radiometer for the full optical path to external radiances. For example the external calibration will, if the radiometer’s design features an external window, include any degradation of that external window. Stirred buckets can have the advantage of being easier to maintain with respect to thermometry and operation. The resulting radiances may also have to be corrected for reflected sky radiation as sea water does not have unity emissivity. This problem can be turned to the user's advantage by using the stirred bucket radiance to provide an “automatic” correction for reflected sky radiation. However, studies using a thermal camera show a >0.1K gradient across a stirred bucket (Jessup 1992) and the cooling effect of wind on the skin temperature can also result in a temperature gradient, even in a well stirred bucket where the skin is being continually broken and renewed (Donlon, 1994). The sky radiation reflected from the water surface in a stirred tank can vary rapidly and by large magnitudes as cloud cover varies.

A solution to the lack of in situ SSST data may be the use of low-cost commercial engineering radiometers that have come onto the market over recent years. Studies have only recently commenced to assess the suitability of these devices in this context (Donlon et al. 1998). The chief drawback of using commercially available radiometers may be the lack of regular, frequent field calibrations provided by a black body or stirred bucket system.

In this chapter a field experiment is described in which a relatively low cost commercial radiometer, the Tasco THI-500L is deployed. A thermal imaging camera, the NEC TH1101 is also deployed. The scheme of instrument calibration and data correction for the Tasco radiometer is presented. A summary of calibration and correction of the NEC TH1101 data performed by Yokoyama and Tanba (1996) in the field experiment is also presented, together with a description of other instrumentation deployed during the experiment. The calibration
results are used to discuss the prospects for greater use of low cost commercial radiometers for *in situ* skin SST measurements. The applications of thermal camera data for study of the variability of small-scale SSST in relation to the skin-bulk temperature deviation and the validation of satellite-measured SSST are also considered.

**2.4 The Mutsu Bay Experiment (MuBEx)**

**2.4.1 Background**

The advent of remote sensing platforms such as ERS and JERS has led to an international initiative aimed at improving expertise and *in situ* data sets for validation and application of this remotely sensed data. Mutsu Bay, northern Honshu Island, Japan was chosen as the site for a series of Japan/UK collaborative campaigns to acquire such an *in situ* data set. These Mutsu Bay Experiment (MuBEx) campaigns were conducted over 3 years. This thesis uses data from the first MuBEx in 1995 (Donlon *et al.*, 1995).

The bay is a maximum of 50km wide but only 10km across at the entrance (Figure 2.1), providing sheltered conditions advantageous for the use of inshore research vessels. In addition, the bay has several meteorological stations along its coast, regular atmospheric radiosonde measurements from the Misawa airbase 50km to the south and contains a network of four instrumented buoys for aquaculture research purposes to which further instrumentation can be added. The UK team who took part in the experiments provided additional equipment, expertise and satellite data from the ERS Along Track Scanning Radiometers ATSR/1 and ATSR/2 for validation purposes. The primary interests of the UK team were validation of the ATSR instrument and investigation of the physical processes occurring at the ocean-atmosphere interface. The MuBEx data has the advantage that it can also be used in studies of SST variability, particularly in the context of the effect of variability in climate modelling.
Figure 2.1: Location of Mutsu Bay Experiment and buoy positions
2.4.2 Instrumentation

Three radiometers were deployed on board the research vessel *Dai-Ni Misago* during the 1995 MuBEx campaign to measure skin SST. The characteristics of two of these instruments are shown in Table 2.1. The thermometry of the third radiometer was found to be unstable, resulting in suspect calibrations and unreliable data.

2.4.2.1 The NEC TH1101 Thermal Infrared Camera

The purpose of the NEC TH1101 thermal infrared camera (TIC) was primarily to provide information on the spatial variation of SSST as well as acting as a validation for the low cost radiometer measurements. The TIC uses a mirror to focus radiation onto a mercury cadmium telluride detector. The detector is cooled by liquid nitrogen. The mirror scans in one direction to build up a scan line of pixels. It is then tilted slightly in the perpendicular direction ready to scan the next line of pixels. This was known as “image scan” mode. A 255×207 pixel image took 8 seconds to acquire. A different “line scan” mode was also used. This operates like a satellite scanner by only detecting one line of pixels and using the motion of the platform to build up consecutive lines of adjacent pixels assuming that the speed of the platform and the scan time are synchronised.

2.4.2.2 The Tasco THI-500L radiometer

The Tasco THI-500L radiometer is another commercially available unit, utilising a thermopile as the detector. The detector is mounted behind a lens that focuses radiative flux onto the detector. The lens is able to collect incident flux from a divergent beam of half angle 4.3°. For a 5m path length this produces a circular field of view of 0.75m diameter. Analogue output is produced by the radiometer, equivalent to 1mV K⁻¹. Thus the output signal is already converted from a flux to a brightness temperature by an internal processor. Further description of the operational use of Tasco THI-500L is given in Donlon et al. (1998). The response functions of the TIC and Tasco radiometers are shown in Figure 2.2 and Figure 2.3. Calculation of SSST from the radiometer data is discussed in section 3.2 in Chapter 3.
Figure 2.2: Response function of the Tasco THI-500L radiometer (data courtesy of Tasco Corporation, Japan)

Figure 2.3: NEC TH1101 response function (courtesy NEC Corporation, Japan)
2.4.2.3 Additional Instrumentation

In addition to assessing the performance of a low cost radiometer for measuring SSST, an aim of MuBEx was to provide data for climate modelling and investigation of the physical processes at the ocean-atmosphere interface. The net ocean-atmosphere energy flux is an important component of both of these areas of study. Therefore, in order for the data set to be of use in these contexts, the parameters required to calculate the net heat flux must be measured. The bulk parameterisations of Smith (1988) and Oberhuber (1988) for obtaining this flux were described in Chapter 1. These require short wave (0.3-2.2μm) solar radiation, wind speed, air temperature, humidity, air pressure and the bulk-skin temperature difference as inputs. These parameters are also essential for the development of satellite SST retrieval algorithms and investigation of the relationship between skin and bulk SST. Meteorological sensors were placed on the Dai-Ni Misago to record these parameters and are shown in Table 2.2. An additional Tasco THI-500L radiometer was deployed to measure the down-welling thermal infrared radiation from the sky. As described in Chapter 1, a small fraction of this is reflected from the sea surface and contributes to the thermal infrared signal measured by the radiometers viewing the sea. This contribution must be removed in order to obtain the correct SSST. A Conductivity Temperature Depth (CTD) probe was deployed to obtain “bulk” sea temperature at 0.5m depth.

<table>
<thead>
<tr>
<th>Instrument name</th>
<th>Detector type</th>
<th>Wavelength range</th>
<th>Resolution</th>
<th>Temperature range</th>
<th>Manufacturer's quoted accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>NEC TH1101</td>
<td>HgCdTe</td>
<td>8-12μm</td>
<td>0.1K</td>
<td>20 to 80°C</td>
<td>± 0.1K</td>
</tr>
<tr>
<td>Tasco THI-500L</td>
<td>Thermopile</td>
<td>8-12μm</td>
<td>0.1K</td>
<td>-50 to 500°C</td>
<td>± 2.0°K</td>
</tr>
</tbody>
</table>
### Table 2.2: Instrumentation deployed on the Dai-Ni Misago

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Instrument</th>
<th>Manufacturers' Quoted Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bulk sea temperature at 0.5m depth</td>
<td>Conductivity/temperature/depth probe (CTD)</td>
<td>±0.01K</td>
</tr>
<tr>
<td>Air temperature</td>
<td>Platinum resistance thermometer</td>
<td>±0.5K</td>
</tr>
<tr>
<td>Relative Humidity</td>
<td>Solid state detector</td>
<td>±1.5% RH</td>
</tr>
<tr>
<td>Short wave (0.3-2.2μm) solar flux</td>
<td>Pyranometer</td>
<td>±10Wm⁻²</td>
</tr>
<tr>
<td>Wind speed and direction</td>
<td>Anemometer (Met station) with vane</td>
<td>±0.1ms⁻¹&lt;br&gt;±3°</td>
</tr>
<tr>
<td>Air pressure</td>
<td>Barometric pressure sensor</td>
<td>±1mb</td>
</tr>
<tr>
<td>Boat speed, direction and location</td>
<td>Global Positioning System (GPS)</td>
<td>Speed ±1ms⁻¹&lt;br&gt;Direction ±10°&lt;br&gt;Location ±10m</td>
</tr>
<tr>
<td>Down-welling thermal infrared sky radiation</td>
<td>Tasco THI-500L radiometer</td>
<td>±0.5K</td>
</tr>
<tr>
<td>Up-welling thermal infrared radiation from sea surface</td>
<td>NEC TH1101 Thermal Infrared Camera, Tasco THI-500L radiometer</td>
<td>NEC ±0.025K&lt;br&gt;Tasco ±0.1K</td>
</tr>
</tbody>
</table>

Figure 2.4 shows the deployment of instrumentation on the *Dai-Ni Misago*. The TIC and Tasco radiometers were mounted on a platform that projected over the bow. This was to keep the angle of incidence to the sea surface as near to the vertical as possible in order to reduce the reflectivity and maximise the emissivity of the sea surface (Masuda *et al.*, 1988). The radiometers had to be inclined at 23° to the vertical in order for the bow to be out of the field of view and to minimise reflection from the ship into the detector. This angle would have been much greater if the radiometers had been mounted directly on the vessel superstructure rather than on the platform. The platform height was 5 metres above the sea surface to minimise sea spray contamination of the instrument. The sky-pointing Tasco was mounted on the platform as one unit with the sea-pointing Tasco. This facilitated easy deployment and removal of both Tasco radiometers for post-cruise calibration on shore. The Tasco unit was covered with reflective foil and insulated using polystyrene packing to minimise the effects of
solar heating, wind cooling and changes in air temperature on the radiometric response. Meteorological sensors were mounted high on the ship to be clear of any effects of the vessel superstructure.

2.4.3 Cruise Strategy

The primary aim of MuBEx was to obtain radiometric “skin” sea surface temperature (SSST) measurements coincident with ERS and NOAA satellite overpasses. This would allow validation of the SST retrieval algorithms applied to the ATSR and AVHRR radiometers on these satellites. Therefore cruises were timed to coincide with the overpass times that happened at a maximum frequency of twice every 3 days (ERS, for 2 satellites) and 4 times every 24 hours (NOAA, for 2 satellites). The research vessel operated within 0.5km of buoy number 6 (see Figure 2.1) near overpass times. This allowed bulk SST and meteorological data to be acquired simultaneously at two locations (the buoy and the boat) within a satellite image pixel in order to investigate intra-pixel spatial bulk SST variation. Buoy 6, near the centre of Mutsu Bay, was chosen because the sea in the area is less susceptible to contamination from aquacultural marine slicks than buoy 4. The area is also calmer than the seas at the mouth of the bay near buoy 1. Due to the inshore nature of the Dai-Ni Misago, the vessel’s deployment was limited to wind speeds of less than Force 5 (approximately 10ms⁻¹). Also, the CTD could not be deployed at these high wind speeds due to the risk of sensor damage from collision with the boat’s hull.

The Dai-Ni Misago was based in Aomori, about 40km from the instrumented Buoy Number 6 (Figure 2.1). This represents a round trip of 3-4 hours at a speed limited to 12 knots to avoid damage to instrumentation from vibrations of the bow platform at higher speeds. The vessel could not be on station for more than 12 hours because of crew shift availability. Therefore, cruises were timed to cover several satellite overpasses over a period of up to 12 hours. In all 11 cruises were undertaken. Table 2.3 shows the scheme of instrumentation deployment for each cruise. Initially, only one Tasco was available and this was used to obtain down-welling sky radiation measurements. The second (sea-pointing) Tasco became available later in the campaign and was deployed for the last 4 cruises. The range of physical conditions encountered during the campaign are shown in Table 2.4. The local and seasonal nature of the dataset can be readily ascertained from this table.
Figure 2.4: The deployment of instrumentation on the R/V Dai-Ni Misago
Table 2.3: Data availability during MuBEx ’95

<table>
<thead>
<tr>
<th>Cruise Date</th>
<th>Tasco radiometer</th>
<th>TIC</th>
<th>Met Sensors</th>
<th>GPS</th>
<th>Buoy no. 6</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sky</td>
<td>Sea</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26/07/95</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>27/07/95</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>29/07/95</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>30/07/95</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>02/08/95</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>03/08/95</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>08/08/95</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>11/08/95</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>12/08/95</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>22/08/95</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>23/08/95</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

Table 2.4: Range of measurements recorded during MuBEx ’95

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Atmospheric Pressure</td>
<td>990mb</td>
<td>1020mb</td>
</tr>
<tr>
<td>Air Temperature</td>
<td>16.1°C</td>
<td>25.7°C</td>
</tr>
<tr>
<td>SSST</td>
<td>19.3°C</td>
<td>26.2°C</td>
</tr>
<tr>
<td>BSST</td>
<td>19.1°C</td>
<td>24.9°C</td>
</tr>
<tr>
<td>Relative Humidity</td>
<td>70%</td>
<td>95%</td>
</tr>
<tr>
<td>Solar Flux</td>
<td>10 Wm^{-2}</td>
<td>1010 Wm^{-2}</td>
</tr>
<tr>
<td>Wind Speed</td>
<td>0.0 ms^{-1}</td>
<td>7.1 ms^{-1}</td>
</tr>
</tbody>
</table>

2.4.4 Calibration strategy

2.4.4.1 Tasco THI-500L radiometer

In order to be easily deployed at a relatively low cost with minimal maintenance, the Tasco radiometers were not subjected to calibration on board the research vessel during each cruise. The relatively short duration of each cruise allowed the radiometers to be calibrated on shore
in the laboratory between each transect. Figure 2.7 shows the black body constructed to enable the on-shore calibration of the radiometers. This was specifically designed to be light and portable to allow it to be stationed near to the dockside so the radiometers could be easily removed from the R/V Dai-Ni Misago, calibrated and re-deployed for the next cruise. The dimensions of the entire assembly, including the surrounding container were approximately 60cm long by 30cm wide by 40cm high.

The calibration target was constructed by Dr. Craig Donlon (University of Southampton) and based on the design of Dr. Ian Barton (CSIRO). The black body cavity is a cylindrical “pot” tapering to a cone with a long axis of 100mm at the interior end. Figure 2.5 shows that this design minimises internal reflections producing near-unity ($\varepsilon>0.999$) emissivity for the Tasco radiometer spot size of 12.25mm used during these calibrations. The cavity was coated with specialist high-emissivity (Figure 2.6) Nextel 101-C10 paint and was constructed from copper to allow good thermal conduction. The black body temperature is controlled by the temperature of the water in the surrounding tank. A digital thermometer, pre-calibrated against a quartz thermometer, was used to record the water temperature to provide a measure of the black body temperature. A stirrer was used to mix the water in order to maximise the thermal homogeneity of the water, and therefore the blackbody. Trials were conducted with several thermometers placed at various locations in the water tank over water temperature ranges from $274K - 308K$. These trials showed good thermal homogeneity (variability $0.05K<\varepsilon$) throughout the water tank while the stirrer was in operation.
Figure 2.5: Emissivities in the 8-12μm band for various black body designs and radiometer spot radii (Nightingale, 1996)

Figure 2.6: Spectral emissivity of Nextel 101-C10 (data from Clarke and Larkin, 1985)
Figure 2.7: The calibration reference black body

The sea-pointing Tasco was prioritised for calibration because SSST measurements are required to be more accurate (±0.1K) than down-welling longwave infrared measurements (±5.0K). Less than 1% of the thermal infrared radiation measured by the sea-pointing Tasco is reflected sky radiation (Masuda et al., 1988). The sea-pointing Tasco was calibrated a total of six times during MuBEx '95. A calibration run consisted of recording the Tasco reading for blackbody temperatures increasing from just above the freezing point of water to 35°C (273K to 308K) over a few hours. This is the range of skin SST that is found around the world.

Six calibration runs were performed on the sea-pointing Tasco radiometer between the cruises during MuBEx. A range of functions was fitted to the resulting calibration curves. In each case a second order polynomial of the form \( y = ax^2 + bx + c \) was found to be the best fit. The minimum change in the Tasco signal for a target temperature of 22°C was 0.045°C for the 36 hours between the 5th and 6th calibrations. This value is just under 50% of the desired accuracy for ATSR validation, suggesting that the maximum desirable time between calibrations is twice this period. That is, 72 hours is the maximum time before drift in instrument response begins to differ significantly from the known interpolated calibration.
with reference to required instrument accuracy. However, this is only the case when the radiometer is deployed during low wind speed (less than $10 \text{ ms}^{-1}$) conditions in inshore waters such as Mutsu Bay. Furthermore, during MuBEx '95 relative humidity was between 70% and 95% and air temperature was always in the range $288 \text{K}$ to $303 \text{K}$. Deployment on ocean-going vessels will most likely result in more varied and rougher conditions leading to more pronounced variation in the radiometer's response and in addition, degradation of the radiometer optics by airborne salt.

The plots of the six values of each polynomial coefficient are shown on Figure 2.8, Figure 2.9 and Figure 2.10. Although the intercept and $\times$ coefficients show a possible periodic behaviour for the last 3-4 calibrations, the low number of data points meant that no periodic curve could be fitted with a high level of confidence. No dependency on the measured physical parameter could be found for this periodicity. Values of air temperature, humidity and pressure corresponding to the calibrations show no periodic behaviour that might coincide with a hypothetical periodic pattern in the coefficient values. Linear interpolation was used to obtain the polynomial coefficients from each graph corresponding to the time and date of each cruise. The interpolated polynomials were then used to correct the cruise data.

![Figure 2.8: Intercept coefficients from polynomials fitted to the six calibrations of the sea-pointing Tasco against the black body calibration source during MuBEx '95](image_url)
Figure 2.9: $x$ coefficients from polynomials fitted to the six calibrations of the sea-pointing Tasco against the black body calibration source during MuBEx '95.

Figure 2.10: $x^2$ coefficients from polynomials fitted to the six calibrations of the sea-pointing Tasco against the black body calibration source during MuBEx '95.
When the Tasco radiometers are deployed during a cruise their response may also be affected over the short and long-term by changes to the local environment. For example, variations in solar flux or air temperature can lead to heating or cooling of the radiometer. In order to try and ascertain the degree of change in calibration of the radiometers due to these two effects, two post-MuBEx '95 experiments were performed using the temperature-controlled rooms at the Southampton Oceanography Centre at Southampton University in the UK.

To quantify the effect of varying air temperature on the Tasco radiometers during cruises, the Tasco radiometers were calibrated at air temperatures of 5, 20 and 35°C using the black body described above. The black body temperature was varied from approximately 0-35 °C for each air temperature. These calibrations were repeated ten months later during the CASOTS I workshop (Donlon et al., 1998) at Southampton Oceanography Centre to investigate whether the variations in response determined from the first experiment were stable. The CASOTS I black body is larger than the target used during the MuBEEx calibrations with a 250mm long cone producing higher emissivity in the 8-12 μm region (Figure 2.5). The Tasco was positioned further away for the CASOTS calibrations, producing a spot size of 25mm.

The two experiments to calibrate the Tasco radiometers at air temperatures of 5, 20 and 35 degrees Centigrade produced 3 calibration curves, one corresponding to each air temperature. A range of functions was fitted to each curve with a second-order polynomial yielding the best fit in each case. An attempt to quantify the difference in each calibration for a typical MuBEx air temperature of 25°C and SSST of 22°C was made by linear interpolation of each polynomial coefficient against temperature. A higher order interpolation with a good degree of confidence was not possible because only 3 points were available for each coefficient. The calibration curve coefficients and the interpolated coefficients for an air temperature of 25°C are shown in Table 2.5.
Table 2.5: Tasco radiometer calibration curve coefficients for different air temperatures.

<table>
<thead>
<tr>
<th>Coef.</th>
<th>First experiment air temperatures</th>
<th>Second experiment air temperatures</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>35°C</td>
<td>20°C</td>
</tr>
<tr>
<td>$x_1^2$</td>
<td>4.54x10^-7</td>
<td>4.33x10^-7</td>
</tr>
<tr>
<td>$x$</td>
<td>1.040</td>
<td>1.038</td>
</tr>
<tr>
<td>$c$</td>
<td>-0.609</td>
<td>-0.477</td>
</tr>
</tbody>
</table>

The interpolated coefficients were used to calculate the difference over the ten-month period in measuring a typical MuBEx SST value of 25°C. This difference was found to be 0.1K, suggesting that the response of the radiometers to changes in air temperature is stable relative to 0.1K measurement accuracy required.

To examine the response of the Tasco radiometers to variations in humidity, the interpolated 25°C calibration curves calculated from the air temperature calibration data (at a low relative humidity of <10%) were compared to the six calibration curves acquired during MuBEx '95 (>70% relative humidity). The coefficients of the low humidity curves were within the maximum and minimum values of coefficients from the MuBEx calibration curves. This implies that significant (>60%) changes in relative humidity do not have a significant effect on the response of the Tasco radiometers.

An experiment to test the response of the Tasco radiometers to changes in solar flux (and therefore heating of the instrument) was also performed at the Southampton Oceanography Centre. The radiometers were calibrated against a CASOTS black body target in a temperature-controlled room with the air temperature set to 298K (a typical MuBEx value) and a black body temperature of 303K (towards the upper limit of SSST around the world). The calibrations were repeated for different values of incident short wave solar flux using 60W daylight simulation light bulbs (produced by Daylight Studios Ltd.) to mimic the solar flux (Figure 1.1). The number of light bulbs in use was varied between 0 and 4 to provide calibrations over a range of solar flux levels. A Kip-Zonen pyranometer similar to the model deployed during MuBEx was used to measure the solar flux incident on the Tasco radiometer. Five calibration points were obtained of solar flux against the difference between the Black Body temperature and the measured Tasco radiometric temperature. The measurements for the solar flux equivalent to 1 daylight simulation bulb were rejected due to anomalous data
from the Tasco radiometer, possibly caused by not allowing the radiometer to stabilise at
the new flux level or from electrical interference.

A curve-fitting algorithm was applied to the calibration points. Two non-linear functions
were found to have chi-squared ($\chi^2$) goodness of fit test values exceeding a 99.5% probability
that the data fits the functions. A linear fit was rejected as it produced a lower chi-squared
probability. The functions are shown in Table 2.6 and Figure 2.11 and Figure 2.12. Errors in
the coefficients were calculated from the measurement errors in the source data. For all
coefficients in Table 2.6 the errors are less than ± 5%.

**Table 2.6: Coefficients for the functions fitted to the solar flux calibrations**

<table>
<thead>
<tr>
<th>BB PRT</th>
<th>a</th>
<th>b</th>
<th>c</th>
<th>$\chi^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exponential</td>
<td>5.89E-02</td>
<td>3.543</td>
<td>-5.62E-02</td>
<td>3.73E-03</td>
</tr>
<tr>
<td>Quadratic</td>
<td>1.174</td>
<td>0.006</td>
<td>1.66E-03</td>
<td>4.27E-03</td>
</tr>
</tbody>
</table>

**Figure 2.11: Exponential curve fitted to solar flux calibration data**
The curves show that the Tasco THI-500L produces warmer measurements as the incident solar flux is increased. The maximum flux levels experienced during the MuBEx '95 project exceeded 1000 Wm$^{-2}$. However, during this time the Tasco assembly was insulated using aluminium foil and polystyrene to minimise the effect of solar heating on the response of the radiometer. Furthermore, the temperature of the insulated assembly was measured using a thermistor located within the assembly. The assembly temperature did not vary from the contemporaneous air temperature by more than 2.0K throughout the MuBEx '95 experiment. During the Southampton experiment, the Tasco was wrapped in aluminium foil in order to assess the effectiveness of insulation on the stability of the Tasco response. The wrapping reduced the warm readings of the Tasco to 0.15K at a solar flux level of 580 Wm$^{-2}$. Calibration of the radiometer at higher levels of solar flux proved difficult, as the number of daylight simulation bulbs that could be physically deployed in the laboratory was restricted to 4. It is inferred that the additional polystyrene insulation used during MuBEx '95 further reduced this warming of the radiometer signal, although this is offset by the greater magnitude of solar flux radiance encountered during MuBEx. Extrapolation of the functions in Table 2.6 shows that peak MuBEx '95 solar flux readings of 1000 Wm$^{-2}$ will produce Tasco “warm” measurements of 1.18K (quadratic fit) or 1.98K (exponential fit). The insulation used during MuBEx will reduce this, although it was not possible to duplicate this in the laboratory in order to quantify the extent to which the effect of solar warming at these levels would be
reduced. If it is assumed that the insulation reduces the warming effect on the radiometer signal to

\[
\frac{\text{the warming observed with the aluminium foil insulation}}{\text{the warming observed without the aluminium foil insulation}} = \frac{0.15K}{0.6K} = 25\%
\]

of the uninsulated value at an incident solar flux of 580Wm\(^{-2}\) then it can be estimated that at an incident solar flux of 1000Wm\(^{-2}\) the insulation will reduce the warming to:

25\% of 1.18K to 1.98K = 0.295K to 0.495K

This value neglects the extra effect of the polystyrene insulation used during MuBEx but returns a maximum error and thus minimum accuracy of 0.5K to be attributed to the MuBEx Tasco THI 500L radiometer data.

These experiments to characterise the response of the Tasco THI-500L radiometer imply that, whilst it is desirable to have on-board calibration of the radiometer during deployment, the use of the radiometers during short (<12hr) cruises in low wind (<10ms\(^{-1}\)) conditions will produce SSST measurements accurate to 0.1K. This holds if the user ignores the effect of heating by solar flux on the calibration during deployment. The radiometers must be calibrated on shore, preferably at the beginning and end of each cruise with a maximum period between calibrations of 72 hours.

Care must be taken to insulate the radiometers from the warming effects of high levels of solar flux. The efficiency of the insulated radiometer assembly should be assessed via laboratory calibration. For the MuBEx '95 data, the accuracy of the Tasco radiometer brightness temperature measurements is reduced to ±0.5K. The effects of solar heating on the insulated MuBEx Tasco assembly could not be tested in the laboratory so the accuracy limit is determined by extrapolation of results for a similarly insulated assembly tested in the laboratory.

### 2.4.4.2 Thermal Infrared Camera Calibration

The NEC TH1101 was subjected to post-MuBEx calibration and noise removal by the collaborating Japanese team (Yokoyama and Tanba, 1996). Noise was found to have been
added to the signal by vibrations from the detector cooling pump switching on and off. This noise was removed by applying a filter to the Fourier transform of the image. The spatial variation of pixel signal across an image was investigated by calibrating the TIC against a curved isothermal black body to allow a near vertical incidence angle for each pixel scan. The image was found to be less sensitive away from the centre of the image in both the row and line directions of the image. The MuBEx '95 data was then corrected by application of the calibration curve to the TIC cruise data.

The TIC uses the difference in signal between an inner reference black body and the outside world to obtain temperature values. The reference blackbody utilised by the TIC is the radiometer chopper, which should ideally remain at a known temperature. The temperature sensor of the chopper was found to have a quantisation of 0.1K resulting in an equivalent fluctuation in the signal. This could not be removed, as the data product from the TIC does not contain the raw chopper temperature. Laboratory calibration experiments by the Japanese team enabled TIC data to be processed to no higher than a ±0.1K accuracy as a consequence of this quantisation.

2.4.5 Other Pre-processing issues

2.4.5.1 Noise on Tasco THI-550L data from the “Yokogowa” logger

Of the six sea-pointing Tasco calibrations during MuBEx, the data from one was recorded using a "Yokogowa" automatic logger. This logger was found to introduce noise onto the signal either due to an impedance mismatch or being improperly earthed. Subsequent calibrations and observations were performed by manually logging the radiometer. This was achieved by recording the logger readout in between sampling times since the logger was found to introduce noise on to the signal only when it recorded data and not on the signal displayed in between sampling points.

This sub-section describes a method developed to retrieve data from automatically logged calibrations and observations by removing the noise and correcting for any offset from a simultaneously acquired manually logged signal. The noise was analysed using three data sets in which both automatic and manual logging took place. Two of these data sets are from observation cruises on 22/08/95 and 23/08/95. The other set is from an experiment on 19/08/95 that used a stirred water tank as a calibration source. As Figure 2.13 shows, the
noisy automatically logged data, \( S_a \), is generally offset from the manually logged data, \( S_m \), with the noise amplitude mostly restricted to well defined limits. It is noticeable that there are seven "anomalous" noise peaks where the logged signal jumps by 0.5-1.0°C as compared to neighbouring values.

![Figure 2.13: Raw automatically logged Tasco data (black) and manually logged interpolated Tasco data (red) for 22/08/95 observation cruise](image)

A Fourier analysis, \( F(f(t)) \), of the automatically logged data was employed to try and identify any strong periodicities in the noise. To remove any trend in the signal from actual variations in the SST, the manually logged signal was removed from the automatically logged data. The sampling interval of the manually logged data was 1 minute and that of the automatically logged data 10 seconds. To maximise the number of data points and minimise the sampling interval to be used in the Fourier analysis, the manually logged data was interpolated to the sampling interval of the automatically logged data. Linear interpolation was applied as a means of extracting intermediate data values without the need to apply more complicated higher order polynomial fits to the data. Maximising the number of data values available for
analysis will minimise the standard deviation of the mean $S^a - S^m$ difference and therefore the statistical error associated with this mean. Minimising the sampling interval allows the Nyquist frequency, the highest frequency that can be sampled by a power spectrum, to be maximised as it is given by:

\[ f_{\text{Nyquist}} = \frac{1}{2T} \]

Where \( T \) is the sampling interval of the data

**Equation 2.1**

To remove a zero frequency peak that would be generated by the general $S^a - S^m$ offset, the mean offset was removed from the signal before applying the transform. A power spectrum of the result from the 22/08/95 observation data is shown in Figure 2.14. This power spectrum calculation is expressed explicitly in the following equation:

\[
P = \frac{1}{N} \left| \mathcal{F}\left[ (S^a - S^m) - (\overline{S^a - S^m}) \right] \mathcal{F}^*\left[ (S^a - S^m) - (\overline{S^a - S^m}) \right] \right|^2
\]

**Equation 2.2**

Where * indicates the complex conjugate and $\overline{S^a - S^m}$ is the mean difference between the automatically logged and linearly interpolated manually logged data.

A fast Fourier transform algorithm in the IDL programming language was used in the calculation of this spectrum. As Figure 2.14 shows, it is clear here that there are no significant periodicities at higher frequencies. The strong peaks at frequencies below 0.01Hz may be due to instrument drift or a trend in the difference between the manually logged and automatically logged signal. The possibility that these peaks were due to variation of the actual SST over the total period of the data set was discounted because the periods corresponding to the peaks (200-1000 seconds) are not equivalent to the total period of the data set (11000 seconds) or modes thereof.

To examine the possibility that there is a trend in the $S^a - S^m$ difference, a 7-point running mean of this value was calculated and its power spectrum plotted. The power spectrum is shown in Figure 2.15. The running mean is calculated by taking the average of the previous three and next three points adjacent to the point being considered. The peaks at frequencies below 0.007 Hz in Figure 2.15 show the low frequency trend in the 7-point running mean of the $S^a -$
$S^m$ difference. This trend appears to have a period of 1000 seconds with modes of this oscillation responsible for the other peaks at frequencies of less than 0.007 Hz. To remove this trend, the 7-point running mean was subtracted from the $S^r-S^m$ value and the power spectrum of the result calculated. This is plotted in Figure 2.16 and clearly shows that the low frequency peaks have been removed, confirming the presence of a periodic trend in the $S^r-S^m$ difference.

A new method to remove the trend in the automatically logged data caused by the actual variation in skin SST was devised. A 7-point running mean of the automatically logged data was used rather than the manually logged data as a measure of trend. The resulting power spectrum is shown in Figure 2.17. The low frequency peaks are not present again confirming that the low frequency peaks in Figure 2.14 are caused by a trend in the $S^r-S^m$ difference. The low power of the higher frequency peaks suggest that there is no significant periodicity to the noise at these frequencies that might be filtered out.

![Figure 2.14: Power spectrum of $S^r - S^m$ for 22/08/95 calibration data using interpolated manually logged data to remove signal trend](image)
Figure 2.15: Power spectrum of a 7-point running mean of the $S^a - S^m$ difference for 22/08/95 data

Figure 2.16: Power spectrum of $S^n - S^m$ minus a 7-point running mean of $S^a - S^m$ for 22/08/95 data
A scheme to find the value of the $S^a - S^m$ offset and remove the noise was devised to allow correction of data that was only recorded by automatic logging. A description of the scheme follows.

A 7 point running mean, $S^a_7$, of $S^a$ was calculated and the difference between this and the interpolated $S^m$ values found. The 7 point running mean was used as it smoothed the logged signal producing a steadier value for the $S^a - S^m$ difference while not contaminating many smoothed values with the anomalous peaks. Figure 2.18 shows a histogram of this difference for one data set. A clear modal peak is visible at about 0.13K. A similar peak was found for each of the three other data sets recorded using the Yokogawa logger. This modal value, as opposed to the mean value, was used as a first approximation to the $S^a - S^m$ difference. This is because the noisy signal has periodic high peaks that would skew the mean value high as shown on Figure 2.13.

In order to remove these peaks and obtain a more representative value of $S^a - S^m$, a test was performed on the data using the modal value of $S^a_7 - S^m$. The standard deviation $\sigma^d$ of the 30
point retrospective running mean of the raw $S^a$ data was calculated as a measure of the normal amplitude of the noise. Values of $S^a - S^m$ were rejected if they satisfied the following criteria:

$$S^a_i - S^m_i > M + 0.05 + 3\sigma^2_i$$

for $i = 0$ to $n$, the number of data points

Where $M$ is the modal difference of $S^a - S^m$ and 0.05K is the measurement error of the manually logged data. This test successfully removes the periodic peaks in the noise as Figure 2.19 shows when compared to Figure 2.13. The mean of the remaining $S^a - S^m$ values was then obtained to produce a final offset value for each of the three data sets. The mean of these final offset values was used in the correction of automatically logged data. The results of this procedure are given in Table 2.7.

![Figure 2.18: Distribution of $S^a - S^m$ difference values for 22/08/95 observation data](image)
Table 2.7: Results of noise offset determination

<table>
<thead>
<tr>
<th>File Date and type</th>
<th>First approximation modal offset</th>
<th>Final mean offset</th>
</tr>
</thead>
<tbody>
<tr>
<td>190895 calibration</td>
<td>0.17</td>
<td>0.138</td>
</tr>
<tr>
<td>220895 observation</td>
<td>0.13</td>
<td>0.134</td>
</tr>
<tr>
<td>230895 observation</td>
<td>0.18</td>
<td>0.152</td>
</tr>
</tbody>
</table>

Final offset value = 0.141 ± 0.012

Figure 2.19: Raw automatically logged Tasco data after removal of anomalous noise peaks (black), and manually logged interpolated Tasco data (red) for 22/08/95 observation cruise

The scheme for processing data sets containing only auto-logged data again took the standard deviation of the 30-point retrospective running mean and used this to test for periodic peaks in the noise. A 7-point running mean of the remaining data was then calculated and the above offset value subtracted. Error propagation throughout the process meant that data retrieved via this method were subject to an additional uncertainty in accuracy of ±0.07K.
2.4.5.2 Sunglint in the Thermal infrared camera data

During the calibration process, the Japanese MuBEx '95 team found the TIC data to have some saturated pixels. By correlating the TIC imagery with simultaneously acquired video camera footage of the sea surface, it was found that the pixel saturation was due to reflected solar radiation in the 8-12\,\mu m infrared wavelengths to which the TIC is sensitive (Figure 2.3). Although the sea surface reflects approximate 2% of incident infrared flux at these wavelengths, if the radiometer is co-incident with the flux directly reflected from the Sun, the possibility of saturation can be readily demonstrated. Reference to Figure 1.1 shows that the Sun, with a temperature of 6000K, emits flux of the order $10^5$ - $10^4$ W m$^{-2}$\,\mu m$^{-1}$ in the 8-12\,\mu m wavelength region of the electromagnetic spectrum. At a angle of incidence of 23° to the vertical, 2% of this radiation is reflected, equating to between 20 and 2 W m$^{-2}$\,\mu m$^{-1}$. For a SSST of 300K the flux being emitted at these wavelengths is of the order of 10 W m$^{-2}$\,\mu m$^{-1}$. The actual energy received by the TIC is the area under the curves of Figure 1.1. Therefore the reflected component of the radiance observed be the TIC can be many times that emitted by the ocean surface and is sufficient to saturate the TIC detector, which has maximum value of 80°C for the brightness temperature it can detect.

This contamination of the radiometric signal has a significant potential to invalidate the statistical tests for SSST variability performed in Chapter 3 and therefore it is necessary to attempt to flag sunglint-contaminated pixels to ensure that they can be removed from the data being tested for variability. The TIC data was received with a 5\sigma threshold sunglint-detection algorithm already implemented by the Japanese MuBEx 95 team. That is, pixels are rejected which have a brightness temperature in the regime:

\[ BT > \bar{x} + 5\sigma \]

Where \(\bar{x}\) is the mean brightness temperature of the image and \(\sigma\) is the standard deviation from the mean. However this procedure did not reject neighbouring pixels to the saturated ones, which may be also contaminated by sunglint but to a lesser degree. The 5\sigma test also does not attempt to follow a “pattern” of contamination across the image.

An enhanced sunglint test was developed that could be applied to the TIC data supplied by the MuBEx team, which had been filtered using the 5\sigma test. The new filter individually tests the eight neighbouring pixels to a pixel found to be saturated by the original 5\sigma test by testing if:
\[ BT > \bar{x} + 2\sigma \]

Pixels with brightness temperatures in this regime were rejected and their eight neighbouring pixels subject to the same test. The limit of \( 2\sigma \) was chosen as \( \bar{x} + 2\sigma \) represents the upper 2.3\% of a normal distribution. This is a compromise between the need to detect and discount pixels that are abnormally “warm” due to sunglint contamination while retaining the “real” variability of the TIC data. Pixels that fail this test and are located near to a pixel determined as saturated by the \( 5\sigma \) thresholding test are more likely to be in the upper 2.3\% of the pixel values than an isolated pixel with a value of \( \bar{x} + 2\sigma \) elsewhere in the image.

The enhanced sunglint test rejected up to 20 times the number of pixels that the \( 5\sigma \) test rejected on MuBEx cruises with clear skies and sunny conditions. The greatest proportion of pixels rejected for any one cruise was 1.90\% and the mean proportion of pixels rejected over all MuBEx TIC data was 0.21\%. The rejection of sunglint-contaminated pixels via this algorithm modified the mean SSST across the affected images by a maximum of 0.1K. The effect of the enhanced sunglint test on the results of the statistical tests applied in Chapter 3 is discussed in sub-section 3.4.4.

2.5 Conclusions

The air temperature calibrations show that commercial Tasco radiometers have a stable response to changes in air temperature, varying by 0.016 degrees over one month for a target temperature of 22°C at an air temperature of 25°C. This variation over one month is less than the 0.1K resolution of the Tasco radiometers.

The varying calibration curve coefficients obtained from data during the MuBEx campaign suggest that the radiometers must be calibrated before and after observation cruises to compensate for instrument drift. The minimum time between MuBEx calibrations was 36 hours giving a difference in the Tasco signal for a 22°C target temperature of 0.045°C. This is close to the 0.05K measurement error of the Tasco radiometers suggesting that 72 hours is the maximum time that should be allowed between calibrations before instrument drift begins to differ significantly from the interpolated calibration curve. This conclusion does not take into consideration short-term variations in the Tasco radiometer calibration that may occur during an observation cruise. Post-MuBEx '95 experiments show that such short-term
variations due to changes in air temperature and humidity are not significant compared to the temperature resolution of the radiometers. However variations in solar flux are found to result in an observed “warming” of the Tasco brightness temperature measurements of 0.4K for a flux of 580Wm$^{-2}$. The functions fitted to the test data for the effect of solar flux on the radiometer response produce an extrapolated heating of 1.18 (quadratic fit) to 1.98K (exponential fit) for solar flux levels of 1000 Wm$^{-2}$. Insulation of the Tasco radiometers reduces this but during MuBEx the Tasco radiometer accuracy must be reduced by this effect to ±0.5K.

The sunglint saturation found in the TIC data has implications for the Tasco data since both are sensitive at similar infrared wavelengths. A simple remedy that rejected Tasco data obtained at the same time as contaminated TIC data was applied to the MuBEx Tasco data set.

The results imply that humidity and air temperature variations during observation cruises will not have a significant effect on the response of the Tasco THI-500L radiometers. The effect of solar flux on the response of the radiometers can be reduced by the use of insulation. However the insulated assembly needs to be characterised in the laboratory to quantify the efficiency of the insulation. For long-term oceanic deployment, an internal calibration arrangement is recommended to enable these commercial radiometers to meet the 0.1K SSST accuracy requirement. Although calibrations from MuBEx ‘95 suggest that the response of the Tasco radiometer is stable within measurement error for up to 72 hours, further study is needed to determine if a maximum required period between calibrations can be justified. A long period between radiometer calibrations is desirable in the context of attempting to reduce the complexity of a radiometer assembly for deployment on "ships of opportunity". However such a recommendation must consider the many factors long-term deployment at sea introduces, which can alter the calibration of the radiometer. Several commercial radiometers have an additional facility that allows air to be pumped through or past the detector aperture. This may enable these radiometers to be cooled and kept free of dust and salt contamination whilst deployed at sea.

An extensive data set from the 1995 MuBEx project has been obtained. This contains sky brightness temperature measurements from the Tasco THI-500L radiometers of sufficient accuracy (± 0.5K) to correct the NEC TH1101 thermal infrared camera brightness temperature data for reflected sky radiation and thus obtain skin SST. The scheme for this is described in the next chapter. The TIC data is accurate enough (±0.1K) to allow investigation
into the behaviour of small-scale SSST variability on satellite validation measurements. A comprehensive set of environmental parameter measurements has also been acquired contemporaneous with the radiometric data, enabling investigation into the effects of these parameters on SSST variability.
3. THE VARIABILITY OF SSST AT SMALL-SCALES

3.1 The need to investigate SSST variability at small-scales

The recent advent of portable, accurate imaging radiometry capable of \textit{in situ} deployment allows the spatial and temporal variability of SSST to be studied at small-scales. This has particular relevance to the validation of satellite-measured SSST where radiometric measurements covering a few metres of the ocean surface are being scaled up to validate satellite pixels with an area of 1km\(^2\) or more. If the radiometric measurement covers an area of high SSST variability in relation to the larger pixel area, it is more likely that the validation measurement will be atypical of the mean SSST across the pixel.

The use of \textit{in situ} radiometry can also be used to explore the variability of small-scale SSST in relation to the local physical parameters that govern \(\Delta T\) to ascertain whether both SSST variability and \(\Delta T\) are driven by the same factors. The variability of small-scale SSST can be compared to that of large-scale SSST as measured by instrumentation mounted on satellite platforms. This has implications for climate modelling. These models often use large-scale SST data but there is little research firstly, as to whether the variability of SSST differs between the small and large-scale data sets or secondly, whether any difference would impact on the model results. In this Chapter and in Chapter 4 the first question is examined by quantifying the variability of SSST for small and large scales respectively.

Minnett (1991) defined ± 2hr time and ± 5km space criteria for validation of the AVHRR 1km\(^2\) resolution BSST product using \textit{in situ} BSST measurements. However SSST variability is generally greater than BSST (Jessup and Hesany, 1996) due to the rapid reaction time of the SSST to changes in wind speed, longwave flux and the other processes which govern SSST. The time and space criteria for validation of high resolution satellite SSST products must therefore be more limiting. Experiments to validate SSST measured by the ATSR instruments on the ERS series satellites have used a one minute average of \textit{in situ} radiometric data. These measurements are acquired at the same time as the satellite overpass and within the area defined by the 1km\(^2\) satellite pixel (Mutlow, 1994; Barton \textit{et al.}, 1995). However, as was described in Chapter 1, this “snap shot” technique represents 0.003\% of the ATSR nadir scan pixel area. Moreover only approximately 0.0001\% of the \textit{in situ} data is acquired at the same time as the ATSR data. If the rapid reaction time of SSST to changes in meteorological
conditions is considered, then the potential error in SSST validation measurements caused by obtaining data in conditions atypical of the 1km² pixel square is concerning. A localised gust of wind, rain squall or variation in the longwave infrared down-welling flux caused by cloud cover variation may all cause an "atypical" SSST measurement to be obtained over one or several minutes.

The averaged SSST (ASST) product from ATSR, consisting of 50 x 50 km cells produced by averaging the high resolution 1 km² pixel product, produces further problems for validation. Donlon and Robinson (1998) used a series of 30 second measurements acquired as the research vessel traversed the ASST cell. These data were used for validation only if they fell within a ± 12 hr time criteria of the satellite overpass. The authors comment that the SSST obtained along the research vessel track may not be fully representative of the cell mean SSST produced by the satellite, especially in areas with large spatial SSST gradients. Furthermore the atmospheric and oceanic conditions which influence SSST may have changed during the time it takes the research vessel to traverse an ASST cell. Donlon and Robinson (1998) reduced these effects by limiting validation measurements to areas with small spatial SSST gradients although this may limit the locations and therefore availability of SSST validation measurements.

In this chapter the measurements from the NEC TH1101 imaging radiometer, together with the rest of the suite of data obtained during the 1995 MuBEx, are used to analyse the small-scale variability of SSST. The objective of the analysis is to produce a recommendation for in situ spatial and temporal tolerances for validation of the ATSR high-resolution product.

Analysis of the small-scale variability of SSST can also yield information on how local physical conditions (wind speed, solar flux etc.) affect the variability of the SSST. In this chapter, two statistical tests for similarity of population are applied to the NEC TH1101 data from MuBEx 1995. The behaviour of the SSST variability is correlated with coincident magnitudes of wind speed, air temperature, solar flux, downwelling longwave flux, BSST, humidity and air pressure. In Chapter 4 these tests are applied to the global ATSR SSST and in situ COADS BSST data sets to examine whether the variability of SSST behaves similarly at large and small-scales.
3.2 Retrieval of in situ SSST from the NEC TH1101

In Chapter 1 it was described how the radiometric signal measured from the sea surface consists of two components: the radiation emitted by the sea surface and the reflected downwelling radiation. In order to extract the true SSST - the temperature of the top 10-50 μm of the ocean - the downwelling sky flux must be measured and used to correct the radiometric signal measured from the ocean surface. In the 1995 MuBEx, the downwelling sky flux was measured using the Tasco THI-500L solid state radiometer. Chapter 2 concluded this instrument produced brightness temperatures (BT) of at least ±0.5K accuracy. Such an accuracy is more than adequate for the purpose of measuring the sky BT for use in calculating SSST. This is because errors in the sky BT do not propagate to the final SSST to the same degree as errors in the sea BT because only approximately 2% of the downwelling flux signal is reflected to contribute to the sea BT. Furthermore, the sky BT is normally much cooler (223 to 293K) than the sea BT, so the magnitude of the downwelling longwave flux incident on the sea surface is much less than the longwave flux emitted by the sea surface.

The SSST is calculated by subtracting the reflected sky radiance from the observed radiance measured by the NEC TH1101 (henceforth referred to as the TIC) to obtain the integrated radiance from the sea skin over the wavelength windows of both radiometers:

\[ W_{SEA} = W_{TIC} - W_{SKY} \]

Equation 3.1

The Tasco THI500-L uses an internal look-up table to convert the detected radiance to an output in brightness temperature (Donlon et al., 1998). The NEC TH1101 imager data was also provided as a calibrated BT. Therefore in order to calculate the reflected sky radiance, a radiance-to-brightness temperature look-up table was constructed using the Planck equation:

\[ W_{SKY} = \int_{\lambda_{lower}}^{\lambda_{upper}} \frac{2hc^2}{\lambda^5 (e^{(hc/\lambda kT_{SKY})} - 1)} (1 - \varepsilon_{\lambda}) \phi_\lambda d\lambda \]

Equation 3.2

Where \( \lambda_{upper} \) and \( \lambda_{lower} \) are the limits of the response of the Tasco THI500-L optics (Figure 2.2); \( h \) is Planck's constant, \( c \) is the speed of light in vacuum; \( k \) is Boltzmann's constant; \( T_{SKY} \) is the radiometric BT of the sky; \( \varepsilon_{\lambda} \) is the emissivity of the sea surface at wavelength \( \lambda \); \( \phi_\lambda \) is
the response of the radiometer at wavelength $\lambda$ (Figure 2.2). The integrations for each $T_{SKY}; W_{SKY}$ look-up pair were calculated over wavelength steps of $10^{-11}$ m to ensure a high accuracy. Values of $\varepsilon_\lambda$ were obtained by interpolation using those of Masuda et al. (1988) for the look angle of $23^\circ$ to the vertical at which the TIC was positioned. Although Masuda et al. (1988) gives wind-dependent emissivities, Watts et al. (1996) conclude that at look angles of less than $40^\circ$, secondary reflections at the sea surface limit the dependency of $\varepsilon$ on surface roughness. The Tasco THI-500L was positioned perpendicular to the angle of the TIC, to measure sky radiation that would be reflected off a planar sea surface into the TIC.

The observed radiance from the sea surface measured by the TIC was also calculated using a look-up table of the Planck equation:

$$W_{TIC} = \int_{\lambda_{lower}}^{\lambda_{upper}} \frac{2hc^2}{\lambda^3 \left( e^{(hc/RTIC)} - 1 \right)} \phi_\lambda d\lambda$$

Equation 3.3

Where $\lambda_{upper}$ and $\lambda_{lower}$ are the limits of the NEC TH1101 filter function (Figure 2.3); $T_{TIC}$ is the brightness temperature of the ocean surface measured by the TIC; $\phi_\lambda$ is the response of the TIC at wavelength $\lambda$ (shown in Figure 2.3) and the other parameters are the same as those in Equation 3.2. There is no emissivity term in Equation 3.3 as we are calculating the total radiance from the sea surface as detected by the TIC. This includes both the emitted and reflected flux components.

Finally, the radiance emitted from the sea surface, $W_{SEA}$, was calculated using the radiances from Equation 3.2 and Equation 3.3 in Equation 3.1. The SSST was then retrieved using a $W_{SEA}; T_{SEA}$ look-up table of the equation:

$$W_{SEA} = \int_{\lambda_{lower}}^{\lambda_{upper}} \frac{2hc^2}{\lambda^3 \left( e^{(hc/RTSEA)} - 1 \right)} \varepsilon_\lambda d\lambda$$

Equation 3.4

Where $T_{SEA}$ is the SSST and $\lambda_{upper}$ and $\lambda_{lower}$ are the maximum limits of both the radiometer response functions. As Figure 2.2 and Figure 2.3 show, the spectral response of both radiometers lies at wavelengths between 8-13 $\mu$m. Atmospheric absorption due to CO$_2$ at $\lambda$
>13.0 μm and O₃ at λ = 9.6 μm is serious for satellite measurements (Zavody et al, 1995), hence the restriction of satellite radiometry to the 10-12μm region. However for the short (= 5m) path lengths between the TIC and ocean surface during MuBEx these effects are not significant. Similarly, absorption due to water vapour over most of the 8-13μm region is weak.

3.3 NEC TH1101 scan geometry

The TIC has an instantaneous field of view (IFOV) subtending an angle of 1.5 mrad to the focus of the instrument. The shape of the “footprint” of the instrument on the sea surface is debatable. Mason (1991) shows that for the ATSR instrument, the “footprint” varies between quadrilateral and elliptical in shape, depending on the look angle. In order to calculate the approximate area of the sea surface that the TIC is viewing, both “footprint” shapes were considered. An IDL software program was written to produce the vertical and horizontal dimensions for each pixel. These depend upon the position of the pixel within the image as the mirror allows swath 30° wide to be scanned.

![Diagram](image)

**Figure 3.1: Along track pixel dimension geometry for the TIC pixel at the centre of the scan line**

If we consider the “along track” length of the pixel, that is, in the direction the research vessel is travelling, for the pixel in the centre of the scan line we have the geometry given in Figure 3.1. The along track pixel length is distance MD. The height of the TIC above the sea surface, distance EEₜ is 5m. Angle MED is 1.5 mrad, the IFOV of the TIC and angle DEEₜ is 23°, the inclination of the TIC to the vertical. This enables MD to be calculated by trigonometric analysis as being $8.85 \times 10^{-3}$ m.
At the end of the scan line, the geometry needed to calculate the along-track pixel dimension is shown in Figure 3.2. In Figure 3.2(a) distance $E_Es$ is the line from the TIC to the point on the sea surface vertically below the TIC. This is the height of the TIUC above the sea surface, 5 m. As in Figure 3.1, $MD$ is the along track pixel dimension for the pixel in the centre of the scan line. $AB$ represents the along track pixel dimension for the pixel at the end of the scan line. Angle $AEB$ is the IFOV of the TIC, 1.5 mrad. In Figure 3.2(b) the front view of the geometry is shown with angle $MEB$ being half of the TIC swath width, $15^\circ$. Trigonometry produces an end-of-scan line along track pixel dimension of $9.16 \times 10^{-3}$ m.

![Diagram of TIC geometry](image)

Figure 3.2: Along track TIC pixel dimension geometry at the end of the scan line as seen from the side (a) and the front (b)

Now we can consider the across-track dimensions of the TIC pixels. Figure 3.3 shows the geometry for the pixel at the centre of the scan line, as seen from an across-track perspective. As in Figure 3.1 and Figure 3.2 the distance $EE_s$ is 5 m, the height of the TIC above the ocean surface with $M$ the mid-point of the scan line. $ZY$ is the across-track dimension of the centre pixel, so angle $ZEY$ is the IFOV of the TIC, 1.5 mrad. As $MEE_s$, the inclination of the TIC to the vertical, $23^\circ$, is also known, $ZY$ can be calculated as $8.15 \times 10^{-3}$ m.
Figure 3.3: Across track pixel dimension geometry for a TIC pixel at the centre of the scan line

The geometry for the across-track dimensions of a pixel at the end of the scan line is shown in Figure 3.4. Points E, Es, and M are as detailed in the previous 3 figures. AQ is the across-track dimension of the end-of-scan-line pixel, therefore angle AEQ is 1.5 mrad, the IFOV of the TIC. Thus AQ is found to be $9.73 \times 10^{-3}$ m.

Software was written in IDL to calculate the along-track and across-track dimensions of all the pixels in a scan line of 255 pixels. The results of this, shown in Figure 3.5, were used to calculate the area of each pixel. The areas for two possible pixel "footprint" shapes were determined: a quadrilateral with the side lengths given by the along- and across-track pixel dimensions or an ellipse with the minor and major axes given by the along- and across-track pixel dimensions. The variation in the computed areas for each pixel with the position of the pixel in the scan line for these two cases is shown in Figure 3.6.

Figure 3.4: Across track pixel dimension geometry for the TIC for a pixel at the end of the scan line
Figure 3.5: Variation of TIC pixel dimensions with the across-track position of the pixel

Figure 3.6: Comparison of TIC pixel footprint area for quadrilateral and elliptical-shaped footprints
As expected, the elliptical case produces smaller areas for each pixel than the quadrilateral case. In the absence of other data about the "footprint" of the TIC IFOV, it is assumed that these two cases represent the upper and lower limits of the pixel area. The IDL program was used to produce an integrated area for a 255×207-pixel TIC image. These were computed to be 3.94m² for quadrilaterally shaped pixels and 3.10m² for elliptical pixels. The true area of the sea surface viewed by the TIC probably lies between these two values.

Analysis of the swath width (30°) and IFOV (1.5 mrad) shows that a 255 pixel line does not represent a spatially continuous scan of the sea surface. That is, 255 × 1.5 mrad ≠ 30°

In fact the integrated 255 pixels scan represents a swath angle of 21.92° and it appears that 8.08° of the ocean skin in the swath is not sampled. Mason (1991) finds that uncorrected ATSR/1 data does not have a uniform sensitivity to radiance across the FOV. Since no geometric correction has been applied to the TIC data it can be assumed that the same is true for the IFOV of the TIC. Therefore it may be the case that the footprint of the TIC is larger than the specifications given by the NEC Corporation. What is most likely is that these specifications represent the full width half maximum (FWHM) sensitivity of the TH1101. This is the area of the sea surface for which the TIC sensitivity is greater than 50% of the maximum sensitivity. So although the area of the sea surface covered under these specifications is discontinuous, it may well be that the TIC detects radiance from the surface surrounding the pixel, but at lower (less than 50% of maximum) sensitivities.

3.4 Statistical tests for the variability of sea surface temperature

3.4.1 Defining “variability”

The need to investigate and quantify the variability of SSST has been described in the earlier sections of this Chapter. This section will outline a methodology for quantifying SSST variability based on the use of statistical tests.

The ocean skin temperature has a fast response time to changes in meteorological parameters (Jessup and Hesany, 1996) particularly in comparison to the response time of BSST. If we consider a TIC SSST image of 3 to 4m² of ocean surface, changes in variability may be manifested as an overall change in the mean SSST of the image, or in a change in the standard deviation from the mean SSST. There are several mathematical procedures available that
produce a test statistic as a measure of whether two or more sets of data are from the same population. The criteria that these procedures used to determine whether these sub-populations are from the same population are based on analysis of variance or comparison of the means of the sub-populations. This is consistent with attempting to discern the degree of "variability" of the SSST as defined above.

Two statistical tests were chosen for application to the TIC SSST data. The analysis of variance F-Test is a powerful test for the similarity of two or more sub-populations. However, the F-Test is a parametric test, which requires that certain assumptions are made about the sub-populations being tested. These assumptions are that the sub-populations have a near-normal distribution and their variances are equal. The implications of violating these assumptions are discussed in detail in sub-section 3.4.2 but the large sample sizes (each image has a maximum of 52785 pixels) of the sub-populations being tested mean that any departures from these assumptions should have negligible effect on the F-Test statistic. Nevertheless, procedures are applied to assess the impact of unequal variances and non-normality and these too are described in sub-section 3.4.2.

A second, non-parametric test was applied to the TIC data as an additional check to see if non-normality and unequal variances of the test populations may be affecting the F-Test results. This second test was the Mann-Whitney U-Test, one of the most powerful distribution-free procedures (i.e. unaffected by the non-normal distributions in the sub-populations being tested). The U-Test requires ranking of the data, as described in sub-section 3.4.3, which is computationally intensive.

The two tests were applied to pairs of TIC images, rather than 3 or more. The aim of this section is to investigate how SSST variability depends on changes in meteorological parameters and therefore heat fluxes. Testing two images to produce a test statistic reflecting the variability between the images is simpler to analyse than attempting to analyse which changes in the several meteorological parameters available are the cause of the variability between three or more images. Testing a larger number of images is also more computationally intensive, particularly with respect to the U-Test. The computational requirements of the U-Test means that only images within the same transect were tested with each other, as between 2,000 to 3,000 TIC images were available for each transect. Therefore, in order to apply both tests to all the unique image pairs in a transect (excluding identities), between $4 \times 10^6$ and $9 \times 10^6$ tests need to be performed per transect. Four
transects from the 1995 MuBEx had co-incident sky radiance data available allowing SSST to be calculated. These transects had a combined total of 10,000 TIC images between them. To test all the unique image pairs would have required $1 \times 10^8$ tests, an increase of a factor of 10 on testing TIC pairs within the same transect.

3.4.2 The F-Test for analysis of variance

3.4.2.1 Definition and calculation of the F-Test

Although the F-Test is described in most texts as an analysis of variance it is in fact a test for a difference between means achieved by using the ratio of two variances. The F-Test assigns a probability that two or more sub-populations are derived from the same global population. That is, any variations within these sub-populations have a high probability of being due to chance variations.

Each TIC image has 52785 pixels, but the sunglint test described in sub-section 2.4.5.2 means that not all the data for each image is available. Therefore the F-Test may be performed on pairs of TIC images that do not have the same number of pixels in each: the sub-populations being tested may have unequal sample sizes. The formulae for deriving the F-Test statistic for sub-populations with unequal sample sizes is a generalised form of the F-Test for populations with equal sample sizes. This allows the F-Test for populations with unequal sample sizes to be used on all TIC image pairs, regardless of whether the image pairs being tested have the same number of valid pixels.

As the F-Test was applied to pairs of TIC images without any other data associated with each pixel, a one-way analysis was required. If other spatial data were available, for example individual measurements of solar flux for each pixel in the image, a two-way analysis could be applied to test whether the spatial solar flux data had an effect on the pixel SSST.

The $F$ distribution is defined as the distribution of the ratio of two independent chi-squared variables; each divided by its degrees of freedom. The F-Test is the ratio of the mean square between the sub-populations to the mean square within the sub-populations. This is used to determine whether two independent estimates of variance derived from these mean squares differ by more than can be reasonably explained on the grounds of errors in the estimates.
There are $M$ different sub-populations being tested, each with $n_i$ elements. The usual unbiased estimate of the variance of each population is:

$$\hat{\sigma}_i^2 = \frac{\sum_j (X_{ij} - \bar{X}_i)^2}{n_i - 1}$$

Equation 3.5

Where $X_{ij}$ are the elements in the sub-population and $\bar{X}_i$ is the mean of the mean of the sub-population. The best overall unbiased estimate of variance is the average of the separate estimates for each sub-population. This is the mean square within, the mean of squared errors within the sub-populations and is expressed as:

$$MS_w = \frac{\sum_i \hat{\sigma}_i^2}{M}$$

Equation 3.6

The sum of the squares within, across all sub-populations is found by summation of Equation 3.5 rearranged for $\sum_j (X_{ij} - \bar{X}_i)^2$, the sum of the squares within each sub-population:

$$SS_w = \sum_i (n_i - 1)\hat{\sigma}_i^2$$

Equation 3.7

However, $(n_i - 1)\hat{\sigma}_i^2 / \sigma^2$ is distributed as $\chi^2$ with $(n_i-1)$ degrees of freedom. The sum of $(n_i-1)$ across all sub-populations is equal to $N$, the total number of elements in all the sub-populations minus $I$. Therefore

$$\sum_i (n_i - 1)\hat{\sigma}_i^2 = \sigma^2 \chi^2_{(N-I)}$$

Equation 3.8

Substituting Equation 3.8 into Equation 3.7 enables $SS_w$ to be expressed in terms of the chi-square distribution:
The mean square between each sub-population is now considered. According to normal distribution theory the means of each sub-population, \( \bar{X}_{ij} \), are normally distributed with the same mean and variance. They can be treated as a random sample of \( M \) observations from the same population. Therefore using the same unbiased estimate of the variance as in Equation 3.5:

\[
\hat{\sigma}_i^2 / n_i = \frac{\sum_i (\bar{X}_{ij} - \bar{X}..)^2}{M - 1}
\]

**Equation 3.10**

Where \( \bar{X}.. \) is the grand mean of all the elements in all the sub-populations. Multiplying Equation 3.10 by \( n_i \) gives an unbiased estimate of the mean square between:

\[
MS_{bet} = n_i (\hat{\sigma}_i^2 / n_i) = \frac{\sum_i n_i (\bar{X}_{ij} - \bar{X}..)^2}{M - 1}
\]

**Equation 3.11**

Which allows calculation of the sum of the squares between:

\[
SS_{bet} = (M - 1)MS_{bet} = \sum_i n_i (\bar{X}_{ij} - \bar{X}..)^2
\]

**Equation 3.12**

Lindman (1974) p.41 shows how it can subsequently be proved that:

\[
SS_{bet} = \sigma^2 \chi^2_{(N-1)}
\]

**Equation 3.13**

Hence the F-Test, a ratio of \( MS_{bet} \) to \( MS_{w} \), can be used to test the hypothesis that two or more sub-populations are drawn from the same overall population by comparing this obtained ration with the \( F \) distribution, the distribution of two independent \( \chi^2 \) variables. The computational formulae for the one-way F-Test for populations with unequal sample sizes are
shown in Table 3.1. A Fortran 90 software program was written to calculate the F-Test statistic for the TIC data using these computational formulae.

Table 3.1: Summary of calculations for one-way F-Test, unequal sample sizes; \( t_{ij} = \sum_{j} X_{ij}; T = \sum_{i} t_{i*} = \sum_{j} \sum_{i} X_{ij} \) (from Lindman, 1974)

<table>
<thead>
<tr>
<th>( RS )</th>
<th>( SS )</th>
<th>( Df )</th>
<th>( MS )</th>
<th>( F )</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td>( T^2/N )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bet (between)</td>
<td>( \sum_{i}(t_{i}^2/n_{i}) )</td>
<td>( RS_{bet}, SS_{m} )</td>
<td>( M-1 )</td>
<td>( SS_{bet}/(M-1) )</td>
</tr>
<tr>
<td>W (within)</td>
<td>( RS_{w}-RS_{bet} )</td>
<td>( N-M )</td>
<td>( SS_{w}(N-M) )</td>
<td></td>
</tr>
<tr>
<td>T</td>
<td>( \sum_{i} \Sigma_{j} X_{ij}^2 )</td>
<td>( RS_{w}, SS_{m} )</td>
<td>( N-1 )</td>
<td></td>
</tr>
</tbody>
</table>

3.4.2.2 Effects of non-normality, unequal variances and unequal sample sizes on the F-Test

The power of the F-Test is defined as the probability that the F-Test can discriminate correctly between sub-populations that are from the same overall population and those that are not. For unequal sample sizes the power of the F-Test is found by calculating the noncentrality parameter:

\[
\phi = \sqrt{\frac{\sum_{i} n_{i} \alpha_{i}^2}{M \sigma^2}}
\]

Equation 3.14

where \( \alpha_{i} \) is the difference between the sub-population mean \( \bar{X}_{ij} \) and the grand mean \( \bar{X}_{..} \).

Charts giving the power of \( F \) as a function of \( \phi \) and the degrees of freedom are produced in Scheffé (1959). The noncentrality parameter \( \phi \) was calculated for each F-Test but the large value of the denominator degrees of freedom (= \( N - 1 \)) meant that the power of the F-Test was never lower than 99%.

The effect of unequal variances is assessed by calculation of \( \lambda \), an estimate of the expected value of \( F \) if the sub-populations are from the same overall populations. \( \lambda \) is given by
\[
\lambda = 1 + \left( \frac{N - 1}{N} \right) \frac{M}{M - 1} \left( \frac{MS_u}{MS_w} - 1 \right)
\]

\textbf{Equation 3.15}

where

\[
MS_u = \left( \frac{1}{M} \right) \sum_i \hat{\sigma}_i^2
\]

\textbf{Equation 3.16}

If \( \lambda \) is close to unity then the F-Test is not greatly affected by unequal variances. Generally for large sample sizes, such as the number of pixels in a TIC image, the ratio between the unequal variances must be very large if the validity of the F-Test is to be affected. Tables giving the probability of errors in \( F \) as a function of unequal values of \( n_i \) and unequal variances are given in Lindman (1974) and Scheffé (1959). The value of \( \lambda \) was calculated for each F-Test and the tables used to find the probability of error in the F-Tests. No cases were found where the error probability was greater than 1%.

The effects of non-normality on the F-Test are examined by Lindman (1974). This analysis concludes that non-zero kurtosis and non-zero skewness in the sub-populations being tested are effectively cancelled out by large values of \( n_i \). Furthermore, it is shown that with only 5 elements in each sub-population the probability of error is not more than 5.3%.

\textbf{3.4.2.3 Results of the F-Tests}

The F-Tests were used to test the null hypothesis that TIC image pairs were from the same population and that variations in SSST between the two images could be accounted for by random fluctuation. Over 12 million F-Tests were performed on unique TIC image pairs from data acquired on 03/08/95 (two transects), 22/08/95 and 23/08/95. The TIC data was sampled at 2 and 5 second intervals during these transects. However co-incident local meteorological data was sampled at intervals varying from 1 second (bulk SST data) to 1 minute (the majority of the meteorological parameters). A polynomial interpolation routine was used to retrieve these meteorological data between observed meteorological measurements. This routine used the 4 observed measurements either side of the time of
acquisition of the TIC image to construct a polynomial curve from which the “missing” meteorological data could be interpolated.

The F-Tests were subject to a threshold of \( F = 7.879 \), equating to a 99.5% confidence of rejecting the null hypothesis. Pairs of TIC images that result in \( F \geq 7.879 \) are therefore “dissimilar”. That is, the differences in SSST between the two images cannot be accounted for by random fluctuations. The absolute difference in meteorological data between each TIC image test pair was calculated in order to assess whether changes in wind speed, solar flux or air temperature etc. were related to the value of the F-Test. The means and standard deviations for the absolute difference in meteorological data for all image pairs that are found to reject the null hypothesis (“dissimilar SSST”), and for images pairs which obey the null hypothesis (“similar SSST”) are shown in Table 3.2. The mean values of the meteorological parameters for these categories are shown in Table 3.3. It is theorised that the value of F-Test is dependent on the variation in the meteorological data between two TIC images, rather than actual values of the meteorological data associated with each individual TIC image. That is, the variation in meteorological data between the two images is one of the factors that govern the variability in SSST. This hypothesis is explored further in this section and later in this Chapter.

The tables show that 0.652% of the tests produce a \( F \) value of \(<7.879\) to confirm the null hypothesis and therefore represent TIC images with “similar” SSSTs. This implies that SSST has a high degree of variability on small-scales. Table 3.2 shows that the mean variation in wind speed, solar flux, air temperature, atmospheric pressure and \( \Delta T \) between images with “dissimilar” SSSTs is greater than that between images with “similar” SSSTs. Therefore, a relatively large change in wind speed (for example) is more likely to result in an increase in the variability of SSST on small-scales. Table 3.3 shows a different behaviour for the means of the meteorological data values (as opposed to the variation of these values between tested images). The mean values for solar flux and air temperature are higher for the “dissimilar” image pairs, but for the other data, the mean values are lower for “dissimilar” image pairs. Therefore, it is an increase in the absolute variation in meteorological data values that is most likely to result in an increase in the variability of SSST on small-scales as measured by the F-Test. The dependency of this measure of SSST variability, the F-Test value, on the absolute variation of the meteorological data between images, is analysed further in the next section.
Table 3.2: Mean meteorological variations for TIC image pairs which produce F values that reject or confirm (shaded) the null hypothesis that the SSSTs in each image are from the same population (99.5% confidence level ($F = 7.879$))

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Global mean</th>
<th>N %</th>
<th>Sum N</th>
<th>Global SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wind Speed</td>
<td>0.6817</td>
<td>99.348</td>
<td>9845067</td>
<td>0.985</td>
</tr>
<tr>
<td>Wind Speed</td>
<td>0.6254</td>
<td>0.652</td>
<td>64581</td>
<td>0.843</td>
</tr>
<tr>
<td>solar flux</td>
<td>295.9214</td>
<td>99.434</td>
<td>12557385</td>
<td>229.600</td>
</tr>
<tr>
<td>solar flux</td>
<td>229.9012</td>
<td>0.566</td>
<td>71436</td>
<td>220.900</td>
</tr>
<tr>
<td>Air temp</td>
<td>0.3227</td>
<td>99.434</td>
<td>12560613</td>
<td>0.417</td>
</tr>
<tr>
<td>Air temp</td>
<td>0.2232</td>
<td>0.566</td>
<td>71436</td>
<td>0.350</td>
</tr>
<tr>
<td>Humidity</td>
<td>1.6472</td>
<td>99.434</td>
<td>12560613</td>
<td>1.497</td>
</tr>
<tr>
<td>Humidity</td>
<td>1.7993</td>
<td>0.566</td>
<td>71436</td>
<td>1.706</td>
</tr>
<tr>
<td>Atmospheric pressure</td>
<td>0.0491</td>
<td>99.663</td>
<td>9065589</td>
<td>0.059</td>
</tr>
<tr>
<td>Atmospheric pressure</td>
<td>0.0464</td>
<td>0.337</td>
<td>30669</td>
<td>0.060</td>
</tr>
<tr>
<td>Sky Temp</td>
<td>3.5606</td>
<td>99.667</td>
<td>6383898</td>
<td>4.394</td>
</tr>
<tr>
<td>Sky Temp</td>
<td>3.475</td>
<td>0.333</td>
<td>21361</td>
<td>5.435</td>
</tr>
<tr>
<td>delta T</td>
<td>0.1507</td>
<td>99.365</td>
<td>9478556</td>
<td>0.124</td>
</tr>
<tr>
<td>delta T</td>
<td>0.0731</td>
<td>0.635</td>
<td>60579</td>
<td>0.075</td>
</tr>
<tr>
<td>Time</td>
<td>2417.455</td>
<td>99.364</td>
<td>9454341</td>
<td>1819.000</td>
</tr>
<tr>
<td>Time</td>
<td>2663.651</td>
<td>0.656</td>
<td>60484</td>
<td>2436.000</td>
</tr>
</tbody>
</table>
Table 3.3: Mean meteorological values for TIC image pairs which produce F values that reject or confirm (shaded) the null hypothesis that the SSSTs in each image are from the same population (99.5% confidence level (F = 7.879))

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean</th>
<th>N %</th>
<th>Sum N</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wind speed</td>
<td>2.9698</td>
<td>99.46</td>
<td>23663676</td>
<td>2.307</td>
</tr>
<tr>
<td>Wind speed</td>
<td>2.9747</td>
<td>0.543</td>
<td>129162</td>
<td>2.279</td>
</tr>
<tr>
<td>Solar flux</td>
<td>922.8412</td>
<td>99.43</td>
<td>25121226</td>
<td>732.50</td>
</tr>
<tr>
<td>Solar flux</td>
<td>732.1258</td>
<td>0.566</td>
<td>142872</td>
<td>627.40</td>
</tr>
<tr>
<td>Air temp</td>
<td>23.4344</td>
<td>99.43</td>
<td>25121226</td>
<td>16.58</td>
</tr>
<tr>
<td>Air temp</td>
<td>23.0157</td>
<td>0.566</td>
<td>142872</td>
<td>16.30</td>
</tr>
<tr>
<td>Humidity</td>
<td>86.2243</td>
<td>99.43</td>
<td>25121226</td>
<td>61.00</td>
</tr>
<tr>
<td>Humidity</td>
<td>86.4651</td>
<td>0.566</td>
<td>142872</td>
<td>61.19</td>
</tr>
<tr>
<td>Atmospheric pressure</td>
<td>787.2365</td>
<td>99.66</td>
<td>18131178</td>
<td>630.50</td>
</tr>
<tr>
<td>Atmospheric pressure</td>
<td>916.5416</td>
<td>0.337</td>
<td>61338</td>
<td>680.30</td>
</tr>
<tr>
<td>Sky Temp</td>
<td>4.9212</td>
<td>99.67</td>
<td>12767796</td>
<td>5.243</td>
</tr>
<tr>
<td>Sky Temp</td>
<td>5.56</td>
<td>0.334</td>
<td>42722</td>
<td>6.229</td>
</tr>
<tr>
<td>delta T</td>
<td>0.4563</td>
<td>99.36</td>
<td>18957112</td>
<td>0.391</td>
</tr>
<tr>
<td>delta T</td>
<td>0.6107</td>
<td>0.635</td>
<td>121158</td>
<td>0.486</td>
</tr>
</tbody>
</table>

3.4.2.4 The dependency of the F-Test on measured physical parameters

One of the aims of this Chapter is to investigate the relationship between the variability of SSST and physical parameters. This is to ascertain whether SSST variability is governed by the same factors as $\Delta T$. It would desirable to obtain an empirical parameterisation of the value of the F-Test in terms of these physical parameters. However, the MuBEx '95 data set is limited in terms of spatial and temporal coverage and any resulting parameterisation might be very specific to these data. Therefore, this analysis will concentrate on completing the initial steps towards a parameterisation by examining the dependency of $F$ on the measured physical parameters. Subsequent research using an in situ data set will then be able to build upon this study to retrieve a generalised parameterisation of $F$.

To investigate the dependency of $F$ on the measured physical parameters, F-Test values were plotted against the variation in each meteorological parameter. The resulting graphs for six of the parameters are shown in Figure 3.11. The large number of points (over 12 million) in each plot results in an unclear graphical representation if a simple scatter plot is used. The
graphs in Figure 3.11 are frequency plots – the data has been sorted into bins with the colour bar showing how many points are in each bin.

A correlation appears to exist between $F$ and the variation in the parameters, wind speed, sky temperature, short wave solar flux, humidity, the bulk-skin temperature difference and time. Initially, a regression algorithm using the Levenberg-Marquandt (Press, 1996) method to minimise $\chi^2$ was written in Fortran 90 to find the curves to fit these data. A first-estimate for the curve was obtained by plotting several “marker” points representing the trend of the relationship. The first-estimate curve fit was found by using the regression module in SPSS on the “marker” plot. A Fortran 90 Levenberg-Marquandt program was written and executed using the first-estimate curve as a starting point. However, this process to minimise $\chi^2$ is valid on data sets with a large number of points, as is confirmed by Thompson (1935).

An alternative approach was adopted to reduce the number of points by averaging the data. The mean value of each meteorological parameter variation was found for 3000 intervals of $F$, equally spaced between the maximum and minimum $F$ values. For example, the mean was found for all values of the variation in wind speed for the range $0 \leq F \leq 1.6 \times 10^4$, then for $1.6 \times 10^4 \leq F \leq 3.2 \times 10^4$ and so on. This produced a 3000 point plot for each meteorological parameter against $F$, to which curves could be fitted using the Levenberg-Marquandt module in SPSS. The $F$ ratio, derived from an analysis of variance table, was used as a goodness of fit test to examine the significance of each relationship. A threshold of 90% for the multiple correlation coefficient, $R^2$, was used to determine whether the polynomial represents a good fit.
Figure 3.7: F-Test statistic versus Meteorological Parameters
The resulting polynomials and the associated goodness of fit values are shown in Table 3.4 and in Figure 3.8 to Figure 3.10. Although the value of $R^2$ for the solar flux regression is 83.8% and therefore below the threshold, the plot of the data in Figure 3.9 clearly suggests a polynomial relationship. The polynomial fit has the greatest deviation from the observed data at $F < 10^8$ but no improvement in $R^2$ was found by trying regressions using higher order polynomial fits or other non-linear functions. It is concluded that the data set shows that there is a polynomial relationship between $F$ and wind speed, solar flux and humidity.

Measurement errors in the original meteorological data were propagated through the calculations by using a weight equivalent to the accuracy of the sensor (Table 2.2) in the calculation to obtain the mean variation in each meteorological parameter for the 3000 plotted F-Test values (Figure 3.8 to Figure 3.10). The standard deviations from each mean variation in these plots were used as weights for the Levenberg-Marquandt curve-fitting algorithm. This yielded a standard deviation, equivalent to the error, for each coefficient given in Table 3.4 of less than 8% of the value of that coefficient.

This analysis has demonstrated that SSST variability, as quantified by the F-Test, is dependent on the variation in wind speed, humidity and solar flux between the tested TIC images. Further analysis to determine the dependency of $F$ on these physical parameters is discussed in sub-section 3.4.5.
Table 3.4: Equations fitted for $F$ against observed variation in physical parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>$R^2$</th>
<th>a</th>
<th>b</th>
<th>c</th>
<th>d</th>
<th>e</th>
<th>f</th>
<th>g</th>
<th>h</th>
<th>i</th>
<th>j</th>
<th>k</th>
<th>l</th>
<th>m</th>
<th>n</th>
</tr>
</thead>
</table>

Note:
1) The form of the polynomial is

$M = a + bF + cF^2 + dF^3 + eF^4 + gF^5 + hF^6 + iF^7 + jF^8 + kF^9 + lF^{10} + mF^{11} + nF^{12} + mF^{13}$

Where $M$ is the variation in the physical parameter and $F$ is the F-Test value.

2) The Multiple Correlation Coefficient, $R^2$, is given by

$R^2 = 1 - \frac{\text{Residual Sum of Squares}}{\text{Corrected Sum of Squares}}$ (Spiegel, 1961)
Figure 3.8: Fitted relationship for variation in wind speed and $F$ value

Figure 3.9: Fitted relationship for variation in solar flux and $F$ value
3.4.3 The Mann-Whitney U-Test

3.4.3.1 Definition and calculation of the U-Test

The procedure applied to quantify SSST variability using the F-Test is now applied using the Mann-Whitney U-Test. This is a distribution-free or non-parametric test with a power >95% of the F-Test, increasing as the sample sizes of the sub-populations increase. The assumptions needed for the F-Test (normally distributed test sub-populations, equal sample sizes and equal variances) are not needed for the U-Test. Thus, the use of the U-Test in this study is primarily as a crosscheck for the F-Test, to ensure that the reliability of the F-Test is not affected by violations of these assumptions.

The U-Test is used to test the hypothesis that two sub-populations are not from the same overall population by examining the means of those sub-populations. Although the U statistic can be calculated directly, for large sub-population sample sizes the sampling distribution of U approaches normal, so the Z score (the deviate of normal distribution) is calculated and compared to the table of Normal deviates.
Initially all elements from both sub-populations are put in a single ranking order. The elements are then split back into their sub-populations and their ranks summed. For example, if there are two sub-populations A and B:

A = 1,3,5,7,9,9  
B = 2,4,4,6,8,10

The elements of A and B are pooled, ranked and then separated with the rank of the elements of either group being summed:

<table>
<thead>
<tr>
<th>A</th>
<th>Rank</th>
<th>B</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>4</td>
<td>4.5</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>4</td>
<td>4.5</td>
</tr>
<tr>
<td>7</td>
<td>8</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>9</td>
<td>10.5</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>9</td>
<td>10.5</td>
<td>10</td>
<td>12</td>
</tr>
</tbody>
</table>

\[ \Sigma R_1 = 39 \quad \Sigma R_2 = 39 \]

The U statistic is then calculated by:

\[ U_1 = n_1n_2 + \frac{1}{2}n_1(n_1 + 1) - R_1 \]

**Equation 3.17**

Where \( n_1 \) and \( n_2 \) are number of elements in the groups A and B respectively. Sometimes this will produce an inflated value of U, which can be detected by calculating \( U_2 \) using:

\[ U_2 = n_1n_2 - U_1 \]

**Equation 3.18**

and taking the lowest U value of the two. Finally, the Z score is given by

\[ Z = \frac{U - \frac{1}{2}n_1n_2}{\sqrt{n_1n_2(n_1 + n_2 + 1)/12}} \]

**Equation 3.19**
As with the F-Test, the Z score is compared to tabulated probabilities that the two sub-populations are dissimilar. If the Z score exceeds the tabulated value for a pre determined probability \( P \) then the two sub-populations have a probability \( \geq P \) of being dissimilar.

The ranking nature of the U-Tests means that calculation of \( U \) with large sample sizes is computationally intensive, as each set of image pairs produce 105570 pixel values to be ranked. Hence, both from the computational perspective and comparison of the power of the tests, the F-Test is the preferable tool to use, provided it is not affected by the parametric limitations of the test. In order to reduce the amount of processing required but still sample the variability across an entire transect, two approaches to processing were undertaken. To sample the variability of the SSST of temporally adjacent images, each image was tested with the 150 following images in the transect. For example, image 1 in a transect was tested with images 2-152, image 2 tested with 3-153 and so on. The second approach sampled the variability of images right across the transect by testing each second image with every fifth succeeding image. Under this procedure image 1 is tested with images 6, 11, 16 and so on, image 3 is tested with images 8, 13, 18 and so on. The U-Test was applied under this scheme via a Fortran 90 program to the TIC data from the 4 transects.

Another computational consideration was the accuracy of the data. The final calibrated TIC product supplied by the Japanese MuBEx 95 team was given to 0.001K, but the accuracy of the data was given as \( \pm 0.1K \) (Tanba, 1996). The data could either be used “as given” and each individual pixel SSST ranked using the 0.001K quantised values or the data could be binned to a lower quantisation that reflected the accuracy of the data. To ascertain the effect of binning the data, the U-Test was applied to a subset of the TIC data using the original 0.001K quantised data and different bin sizes up to a maximum of 0.5K. It was found that the resulting Z values corresponded to confidence levels which varied by less that 10% between the 0.001K quantised data and data binned into 0.1K steps. However increasing the bin size to 0.5K produced a 40% deviation in confidence levels from those given by the 0.001K quantisation. This suggests that the detail of the variability in SSST over area represented by the TIC image is reduced to a level which severely affects the ability of the U-Test to act as a measure of SSST variability. It was therefore concluded that the TIC SSST data should be binned into 0.1K steps prior to being subjected to the ranking module of the U-Test. This binning routine was added to the Fortran 90 program written to perform the U-Test.
3.4.3.2 Results of the U-Tests

Over 2.4 million TIC image pairs were subjected to a U-Test in order to test the null hypothesis that any differences in the two populations are due to random variations in the images. As with the F-Test the U-Test pairs were subject to a Z threshold representing a 99.5% confidence of rejecting the null hypothesis. This is equivalent to a Z value of 2.8. U-Tests producing a Z value of ≥2.8 represents a rejection of the null hypothesis, that is the two TIC images tested are dissimilar. U-Tests with a Z value of <2.8 represent a confirmation of the null hypothesis that any differences between the two TIC images are due to random fluctuation. The variation in meteorological parameters for each image pair was calculated and the means of these variations for each category are shown in Table 3.5 and Table 3.6. The mean values of each parameter obtained are shown in Table 3.7 and Table 3.8.

The tables show that 0.891% of the 150-cycle U-Tests produce a Z value of <2.8 and so may be regarded as being from the same population. This proportion declines to 0.365% for the 5/2-cycle U-Tests. This confirms the conclusions from the F-Test analysis that the SSST shows great variability on small-scales. Table 3.5 shows that, for the 150-cycle scheme, the mean variation in all data between images with “dissimilar” SSSTs is greater than that between images with “similar” SSSTs. Table 3.6 shows that for the 5/2-cycle scheme, the same is true for the mean variation in wind speed, solar flux, air temperature, atmospheric pressure and ΔT. This agrees with the F-Test results given in Table 3.2. The 5/2-cycle scheme is more representative than the 150-cycle scheme of all TIC image pair combinations in a transect being tested against each other. Therefore, it is to be expected that the 5/2-cycle scheme will produce similar results to those of the F-Test procedure, which tested all TIC images combinations in each transect. Again agreeing with the F-Test results of Table 3.3, Table 3.7 and Table 3.8 show a different behaviour for the means of the meteorological data. Fewer meteorological parameters have higher mean values for the “dissimilar” image pairs. Therefore the conclusions of the F-Test are confirmed by the U-Test: it is an increase in the absolute variation in meteorological data values that is most likely to result in an increase in the variability of SSST on small-scales as measured by the U-Test.

The agreement between the F-Test and U-Test supports the assertion that these tests are producing similar results and both reject the null hypothesis for the same pairs of TIC images. A direct comparison of the F and U-Tests is undertaken in sub-section 3.4.4.7. The dependency of the SSST variability, as quantified by the U-Test, on the absolute variation of the meteorological data between images is analysed further in the next section.
Table 3.5: Mean meteorological variations for TIC pairs which fail & pass (shaded) the U-Test 99.5% confidence of populations being dissimilar (Z = 2.8) (150 cycle scheme)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>N %</th>
<th>Sum N</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wind speed</td>
<td>0.4361</td>
<td>99.109</td>
<td>1391346</td>
<td>0.788</td>
</tr>
<tr>
<td>Wind speed</td>
<td>0.3132</td>
<td>0.891</td>
<td>12504</td>
<td>0.641</td>
</tr>
<tr>
<td>solar flux</td>
<td>122.6613</td>
<td>99.109</td>
<td>1391107</td>
<td>143.50</td>
</tr>
<tr>
<td>solar flux</td>
<td>70.0514</td>
<td>0.891</td>
<td>12502</td>
<td>105.80</td>
</tr>
<tr>
<td>Air temp</td>
<td>0.1095</td>
<td>99.109</td>
<td>1391346</td>
<td>0.297</td>
</tr>
<tr>
<td>Air temp</td>
<td>0.0804</td>
<td>0.891</td>
<td>12504</td>
<td>0.268</td>
</tr>
<tr>
<td>Humidity</td>
<td>0.7148</td>
<td>99.109</td>
<td>1391346</td>
<td>0.857</td>
</tr>
<tr>
<td>Humidity</td>
<td>0.6528</td>
<td>0.891</td>
<td>12504</td>
<td>0.888</td>
</tr>
<tr>
<td>Atmos. Pressure</td>
<td>0.0322</td>
<td>99.371</td>
<td>715918</td>
<td>0.049</td>
</tr>
<tr>
<td>Atmos. Pressure</td>
<td>0.025</td>
<td>0.629</td>
<td>4532</td>
<td>0.044</td>
</tr>
<tr>
<td>Sky Temp</td>
<td>1.8682</td>
<td>99.364</td>
<td>721835</td>
<td>3.846</td>
</tr>
<tr>
<td>Sky Temp</td>
<td>1.6398</td>
<td>0.636</td>
<td>4620</td>
<td>4.452</td>
</tr>
<tr>
<td>delta T</td>
<td>0.1005</td>
<td>99.099</td>
<td>1346972</td>
<td>0.108</td>
</tr>
<tr>
<td>delta T</td>
<td>0.0459</td>
<td>0.901</td>
<td>12253</td>
<td>0.066</td>
</tr>
<tr>
<td>Time</td>
<td>258.9523</td>
<td>99.108</td>
<td>1394598</td>
<td>163.10</td>
</tr>
<tr>
<td>Time</td>
<td>234.6676</td>
<td>0.892</td>
<td>12552</td>
<td>188.20</td>
</tr>
</tbody>
</table>
Table 3.6: Mean meteorological variations for TIC image pairs which fail & pass (shaded) the U-Test 99.5% confidence level for dissimilar populations \((Z = 2.8)\) \((5/2\) cycle processing scheme) 

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>N %</th>
<th>Sum N</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wind speed</td>
<td>0.689</td>
<td>99.635</td>
<td>1182254</td>
<td>0.932</td>
</tr>
<tr>
<td>Wind speed</td>
<td>0.6308</td>
<td>0.365</td>
<td>4330</td>
<td>0.888</td>
</tr>
<tr>
<td>solar flux</td>
<td>294.3827</td>
<td>99.635</td>
<td>1182110</td>
<td>241.70</td>
</tr>
<tr>
<td>solar flux</td>
<td>210.8093</td>
<td>0.365</td>
<td>4330</td>
<td>204.50</td>
</tr>
<tr>
<td>Air temp</td>
<td>0.3228</td>
<td>99.635</td>
<td>1182254</td>
<td>0.418</td>
</tr>
<tr>
<td>Air temp</td>
<td>0.2473</td>
<td>0.365</td>
<td>4330</td>
<td>0.373</td>
</tr>
<tr>
<td>Humidity</td>
<td>1.6632</td>
<td>99.635</td>
<td>1182254</td>
<td>1.502</td>
</tr>
<tr>
<td>Humidity</td>
<td>1.9432</td>
<td>0.365</td>
<td>4330</td>
<td>1.812</td>
</tr>
<tr>
<td>Atmos Pressure</td>
<td>0.0634</td>
<td>99.777</td>
<td>632979</td>
<td>0.072</td>
</tr>
<tr>
<td>Atmos Pressure</td>
<td>0.053</td>
<td>0.223</td>
<td>1416</td>
<td>0.07</td>
</tr>
<tr>
<td>Sky Temp</td>
<td>3.5628</td>
<td>99.775</td>
<td>637372</td>
<td>4.467</td>
</tr>
<tr>
<td>Sky Temp</td>
<td>3.4399</td>
<td>0.225</td>
<td>1440</td>
<td>5.723</td>
</tr>
<tr>
<td>delta T</td>
<td>0.2366</td>
<td>99.627</td>
<td>1133986</td>
<td>0.284</td>
</tr>
<tr>
<td>delta T</td>
<td>0.0749</td>
<td>0.373</td>
<td>4243</td>
<td>0.078</td>
</tr>
<tr>
<td>Time</td>
<td>2778.461</td>
<td>99.633</td>
<td>1187433</td>
<td>2114.00</td>
</tr>
<tr>
<td>Time</td>
<td>3488.361</td>
<td>0.367</td>
<td>4374</td>
<td>2495.00</td>
</tr>
</tbody>
</table>
Table 3.7: Mean meteorological values for TIC image pairs which fail & pass (shaded) the U-Test 99.5% confidence level for dissimilar populations ($Z = 2.8$) (150 cycle processing scheme)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>N %</th>
<th>Sum N</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wind speed</td>
<td>2.9309</td>
<td>99.11</td>
<td>2782692</td>
<td>2.352</td>
</tr>
<tr>
<td>Wind speed</td>
<td>2.8358</td>
<td>0.891</td>
<td>25008</td>
<td>2.248</td>
</tr>
<tr>
<td>solar flux</td>
<td>901.3129</td>
<td>99.11</td>
<td>2782692</td>
<td>769.60</td>
</tr>
<tr>
<td>solar flux</td>
<td>763.2893</td>
<td>0.891</td>
<td>25008</td>
<td>673.60</td>
</tr>
<tr>
<td>Air temp</td>
<td>23.3701</td>
<td>99.11</td>
<td>2782692</td>
<td>16.54</td>
</tr>
<tr>
<td>Air temp</td>
<td>23.1157</td>
<td>0.891</td>
<td>25008</td>
<td>16.36</td>
</tr>
<tr>
<td>Humidity</td>
<td>86.3247</td>
<td>99.11</td>
<td>2782692</td>
<td>61.09</td>
</tr>
<tr>
<td>Humidity</td>
<td>86.333</td>
<td>0.891</td>
<td>25008</td>
<td>61.10</td>
</tr>
<tr>
<td>Atmos Pressure</td>
<td>1009.81</td>
<td>99.37</td>
<td>1431836</td>
<td>714.00</td>
</tr>
<tr>
<td>Atmos Pressure</td>
<td>1009.825</td>
<td>0.629</td>
<td>9064</td>
<td>714.00</td>
</tr>
<tr>
<td>Sky Temp</td>
<td>4.8513</td>
<td>99.36</td>
<td>1443670</td>
<td>5.894</td>
</tr>
<tr>
<td>Sky Temp</td>
<td>5.6961</td>
<td>0.636</td>
<td>9240</td>
<td>6.735</td>
</tr>
<tr>
<td>delta T</td>
<td>0.3661</td>
<td>99.10</td>
<td>2693944</td>
<td>0.548</td>
</tr>
<tr>
<td>delta T</td>
<td>0.4838</td>
<td>0.902</td>
<td>24506</td>
<td>0.512</td>
</tr>
</tbody>
</table>
Table 3.8: Mean meteorological values for TIC image pairs which fail & pass (shaded) the U-Test 99.5% confidence level for dissimilar populations (Z = 2.8) (5/2 cycle processing scheme)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>N %</th>
<th>Sum N</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wind speed</td>
<td>2.9695</td>
<td>99.64</td>
<td>2364508</td>
<td>2.308</td>
</tr>
<tr>
<td>Wind speed</td>
<td>2.9847</td>
<td>0.365</td>
<td>8660</td>
<td>2.315</td>
</tr>
<tr>
<td>Solar flux</td>
<td>901.0795</td>
<td>99.64</td>
<td>2364508</td>
<td>732.20</td>
</tr>
<tr>
<td>Solar flux</td>
<td>683.0472</td>
<td>0.365</td>
<td>8660</td>
<td>611.30</td>
</tr>
<tr>
<td>Air temp</td>
<td>23.4064</td>
<td>99.64</td>
<td>2364508</td>
<td>16.56</td>
</tr>
<tr>
<td>Air temp</td>
<td>22.9641</td>
<td>0.365</td>
<td>8660</td>
<td>16.25</td>
</tr>
<tr>
<td>Humidity</td>
<td>86.0797</td>
<td>99.64</td>
<td>2364508</td>
<td>60.90</td>
</tr>
<tr>
<td>Humidity</td>
<td>86.2825</td>
<td>0.365</td>
<td>8660</td>
<td>61.05</td>
</tr>
<tr>
<td>Atmos. Pressure</td>
<td>1009.813</td>
<td>99.78</td>
<td>1265958</td>
<td>714.00</td>
</tr>
<tr>
<td>Atmos. Pressure</td>
<td>1009.824</td>
<td>0.223</td>
<td>2832</td>
<td>713.90</td>
</tr>
<tr>
<td>Sky Temp</td>
<td>4.9179</td>
<td>99.77</td>
<td>1274744</td>
<td>5.518</td>
</tr>
<tr>
<td>Sky Temp</td>
<td>5.4069</td>
<td>0.225</td>
<td>2880</td>
<td>6.778</td>
</tr>
<tr>
<td>delta T</td>
<td>0.376</td>
<td>99.63</td>
<td>2267972</td>
<td>0.447</td>
</tr>
<tr>
<td>delta T</td>
<td>0.5795</td>
<td>0.373</td>
<td>8486</td>
<td>0.504</td>
</tr>
</tbody>
</table>

3.4.3.3 The dependency of U on the measured physical parameters

To investigate the dependency of U on the measured physical parameters, Z values were plotted against the variation in each meteorological parameter. The resulting graphs for six of the parameters are shown in Figure 3.11. As with the F-Test, the large number of points (over 2 million) in each plot results in an unclear graphical representation if a simple scatter plot is used. The graphs in Figure 3.11 are frequency plots – the data has been sorted into bins with the colour bar showing how many points are in each bin.

The graphs in Figure 3.11 show that a relationship appears to exist between Z and the variation in the parameters, wind speed, sky temperature, short wave solar flux, relative humidity, $\Delta T$ and time. The same procedure as applied to the F-Test, averaging the data to reduce the number of data points, was applied. The mean value of each meteorological parameter variation was found for 3000 intervals of U, equally spaced between the maximum and minimum Z values. This produced a 3000 point plot for each meteorological parameter against Z, to which curves could be fitted using the Levenberg-Marquandt module in SPSS.
The resulting polynomials and the associated goodness of fit values are shown in Table 3.9 and Figure 3.12 to Figure 3.18. At high values of $Z$ the observed data differ significantly from the fitted curves. The maximum $Z$ values produced by the U-Test did not exceed 287 and these represent less than 1% of the total tests. To be confident that the fitted curve represents an accurate regression of the data, both in qualitative and quantitative terms, it proved necessary to exclude high values of $Z$ prior to the regression. The limitations with respect to the $Z$ value of the validity of the fitted curves are given in the final column of Table 3.9.

Table 3.9 shows only two fitted polynomials with $R^2$ values greater than 0.9. These are the relationships for solar flux and $\Delta T$. However, reference to the respective plots shows that the other polynomials are qualitatively a good fit. The low value of $R^2$ for the polynomials fitted to wind speed, relative humidity and sky temperature is due to the larger spread of observed data points about the fitted curve. For air temperature and time between images, the fitted functions deviate from the observed data for specific regimes of the variation in the physical parameter. For the air temperature relationship (Figure 3.14), the fitted function deviates significantly when the variation in air temperature exceeds 0.36K. Figure 3.18 shows that the fitted function deviates from the observed data when $t$, the time between images, is less than 320 seconds. The $R^2$ value for this fit may also be reduced by the fitted function tending to one side of the spread of observed values when $t > 450$ seconds. However the fit improves for $t > 600$ seconds.

Measurement errors in the original meteorological data were propagated through the calculations by using a weight equivalent to the accuracy of the sensor (Table 2.2) in the calculation to obtain the mean variation in each meteorological parameter for the 3000 plotted $Z$ scores (Figure 3.12 to Figure 3.18). The standard deviations from each mean variation in these plots were used as weights for the Levenberg-Marquardt curve-fitting algorithm. This yielded a standard deviation, equivalent to the error, for each coefficient given in Table 3.9 of less than 5% of the value of that coefficient.

This analysis has demonstrated that SSST variability, as quantified by the U-Test, is dependent on the variation in wind speed, air temperature, humidity, time, sky temperature, $\Delta T$ and solar flux between the tested TIC images. Further analysis to determine the dependency of $U$ on these physical parameters is discussed in section 3.4.5.
Table 3.9: Equations fitted for U against observed variation in physical parameters

<table>
<thead>
<tr>
<th>Physical parameter</th>
<th>Fitted Equation</th>
<th>$R^2$</th>
<th>Validity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wind Speed (WS)</td>
<td>$U = 280.56 - \exp[-7.88 \log_e(WS)]$</td>
<td>0.150</td>
<td>$U &lt; 280.57$</td>
</tr>
<tr>
<td>Solar Flux (SF)</td>
<td>$U = -1360.3 + 296.78 \log_e(SF)$</td>
<td>0.944</td>
<td>$U &lt; 276.66$</td>
</tr>
<tr>
<td>Air Temperature (AT)</td>
<td>$U = 280.25 - \exp[1.36/AT]$</td>
<td>0.290</td>
<td>$U &lt; 280.25$</td>
</tr>
<tr>
<td>Relative Humidity (RH)</td>
<td>$U = -6.0 \times 10^{-8} RH^3 + 1.7 \times 10^{-3} RH^2 + 2 \times 10^{-2} RH + 1.09$</td>
<td>0.597</td>
<td>$U &lt; 279.06$</td>
</tr>
<tr>
<td>Sky Temperature (ST)</td>
<td>$U = -14.65 ST^3 + 108.95 ST^2 + 121.91 ST$</td>
<td>0.553</td>
<td>$U &lt; 284.25$</td>
</tr>
<tr>
<td>$\Delta T$</td>
<td>$U = -9057.9 \Delta T^2 + 4250.75 \Delta T - 186.81$</td>
<td>0.988</td>
<td>$U &lt; 276.86$</td>
</tr>
<tr>
<td>Time (T)</td>
<td>$U = [(280.36)^2 + 17.62 (0.9807^t)]^t$</td>
<td>0.769</td>
<td>$U &lt; 280.36$</td>
</tr>
</tbody>
</table>
Figure 3.11: U-Test statistic versus Meteorological Parameters
Figure 3.12: Fitted relationship for U-Test Statistic and variation in wind speed

Figure 3.13: Fitted relationship for U-Test Statistic and variation in Solar Flux
Figure 3.14: Fitted relationship for U-Test Statistic and variation in Air Temperature

Figure 3.15: Fitted relationship for U-Test Statistic and variation in Relative Humidity
Figure 3.16: Fitted relationship for U-Test Statistic and variation in Sky Temperature

Figure 3.17: Fitted relationship for U-Test Statistic and variation in $\Delta T$
3.4.4 The effect of the sunglint-detection algorithm on F and U-test results

The identification and removal by two detection algorithms of saturated pixels in the TIC data, caused by reflected solar infrared flux was described in sub-section 2.4.5.2. This section aims to ascertain whether the algorithms will produce differing results when the F and U-Tests for variability are applied to the data with the saturated pixels removed. A subset of 100 TIC image pairs containing saturated pixels were subjected to the F and U-Tests. Three versions of the subset of TIC data were processed: "unprocessed" data with the sunglint-saturated pixels present; data with the "standard" 5σ thresholding algorithm applied; data with the enhanced detection algorithm applied. The mean, maximum and minimum deviations of the resultant F and U-Test values from those for "unprocessed" TIC data are shown in Table 3.10.
Table 3.10: Change in F and U-Test values after application of sunglint-detection algorithms on TIC data

<table>
<thead>
<tr>
<th>Change compared to unprocessed data</th>
<th>Standard 5σ Algorithm</th>
<th>Enhanced Algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td>F-Test (Z value)</td>
<td>Mean change</td>
<td>3.23×10⁶</td>
</tr>
<tr>
<td></td>
<td>Minimum change</td>
<td>1.26×10⁵</td>
</tr>
<tr>
<td></td>
<td>Maximum change</td>
<td>1.78×10⁷</td>
</tr>
<tr>
<td>U-Test (Z value)</td>
<td>40.7</td>
<td></td>
</tr>
<tr>
<td>F-Test (Z value)</td>
<td>6.53×10⁶</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.5×10⁷</td>
<td></td>
</tr>
<tr>
<td>U-Test (Z value)</td>
<td>75.8</td>
<td></td>
</tr>
<tr>
<td></td>
<td>10.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>210.5</td>
<td></td>
</tr>
</tbody>
</table>

The results show that the standard sunglint algorithm has a marked effect on the F and U-Test values. If one considers that the P(.995) value for $F$ is 7.879 then it can be seen that changes to the value of $F$ in the range $1.26\times10^5$ to $1.78\times10^7$ as a result of removing saturated pixels with the standard algorithm are very significant. The same is true with the $Z$ scores from the U-Test. The P(.995) value for $Z$ is 2.8 so changes in $Z$ of the range 3.5 to 157.2 after application of the standard algorithm are again very significant. The results from the use of the enhanced algorithm show even greater differences from those using the "unprocessed" data. Therefore the use of the enhanced saturation detection algorithm is essential to ensuring that the TIC data contains the "true" variability of the SSST and has not been biased by contamination with anomalous pixels saturated by reflected solar flux.

3.4.5 The form of the dependency of $F$ and $U$-Tests on observed physical parameters

This study has now quantified SSST variability on small-scales using the statistical F and U-Tests. The limited spatial and temporal coverage of the MuBEx '95 has led the author to conclude that attempting to derive an empirical parameterisation in terms of the physical parameter measured during MuBEx '95 would not be valid unless an in situ data set with better coverage could be utilised. However, finding a more qualitative general dependency of SSST variability on these physical factors is unlikely to be affected as much by the limitations of the available data. Moreover, such an analysis would form a foundation on which future research using data that are more comprehensive could develop.

To determine the form of the dependency of $F$ or $U$ in terms of the observed physical parameters, it is necessary to analyse whether $F$ or $U$ is a sum of the function of each
parameter, a function of all the parameters, or some combination of both. That is, if our
parameters are $A$, $B$ and $C$ then for $F$ the three possibilities are:

$$F = f(A) + f(B) + f(C)$$

**Equation 3.20** (sum of a function of each parameter)

$$F = f(A) + f(B, C) \text{ or } F = f(A, B) + f(C) \text{ or } F = f(A, C) + f(B)$$

**Equation 3.21** (combination)

$$F = f(A, B, C)$$

**Equation 3.22** (a function of all parameters)

To investigate which of these relationships is true for the data, it must be ascertained whether
each meteorological parameter is dependent on any other measured meteorological
parameters. This was achieved by performing a curve fitting procedure using the SPSS
software on each combination of 2 meteorological parameters. A combination of intuitive
curve fitting and the curve estimation module, which attempts to fit 11 types of function to the
data, were used to fit the data. If necessary this was complemented by fitting user-defined
non-linear curves using a separate SPSS module. The $F$ ratio was used as a goodness of fit
test. A threshold of 90% for the value of the multiple correlation coefficient $R^2$ was used to
determine whether there is a dependency between variables. Table 3.11 shows the matrix of
whether a relationship could be found for these tests under the above criteria.

<table>
<thead>
<tr>
<th>Wind Speed</th>
<th>Solar Flux</th>
<th>Air Temp.</th>
<th>Relative Humidity</th>
<th>Sky Temp.</th>
<th>$\Delta T$</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wind Speed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Solar Flux</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Air Temp.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Relative Humidity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sky Temp.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Delta T$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
In sub-section 3.4.2.4 it was found that $F$ is related to the variation in wind speed, solar flux and relative humidity. The analysis of the relationships between the variation in the measured physical parameters, summarised in Table 3.11, shows that the variation in wind speed is dependent on the variation in solar flux. However, the analysis shows that the variation in relative humidity is dependent on the variation in solar flux. Referring to the conclusions from sub-section 3.4.2.4 and Equation 3.20 to Equation 3.21 this leads to the conclusion that the dependency of the F-Test statistic on the variation in the measured physical parameters takes the form of Equation 3.22:

$$F = f(WS, SF, RH)$$

Equation 3.23

Where $WS$, $SF$ and $RH$ are the variations between TIC images of wind speed, solar flux and relative humidity respectively.

In sub-section 3.4.3.3 it was found that the $Z$ value given by the U-Test is dependent on the variation in wind speed, solar flux, sky temperature, relative humidity, $\Delta T$ and time between TIC images. Table 3.11 shows that all these factors are inter-dependent: the variation in wind speed is dependent on the variation in all of these factors except for relative humidity. However, the variation in relative humidity is dependent on the variation in solar flux. Referring to the conclusions of sub-section 3.4.3.3 and Equation 3.20 to Equation 3.22 this leads to the conclusion that the dependency of the $Z$ value, derived from the U-Test, on the variation in the measured physical parameters takes the form of Equation 3.22:

$$Z = f(WS, SF, AT, RH, ST, \Delta T, t)$$

Equation 3.24

Where $WS$, $SF$, $AT$, $RH$, $ST$, and $\Delta T$ are the variations between TIC images of wind speed, solar flux, air temperature, relative humidity, sky temperature and the bulk-skin temperature difference respectively. The time between TIC images is given by $t$. 
Equation 3.23, Equation 3.24 and the regressions in sub-sections 3.4.2.4 and 3.4.3.3 show that the U and F-Tests are dependent on a limited number of meteorological factors. However, by examining macro- and mesoscale climate models (for example: Ma et al., 1994, Dewitte and Perigaud, 1996, Stockdale, 1992) it is clear that these meteorological factors are dependent on each other. The limitations of the MuBEx 95 data set, in terms of the range of conditions measured and the geographical coverage of data, are most likely to account for other meteorological and physical parameters not appearing to form part of the parameterisations given in Equation 3.23 and Equation 3.24.

3.4.6 Is SSST variability governed by the same factors as air-sea heat fluxes and $\Delta T$?

The previous section derived the forms of the dependencies of the F-Test statistic and the Z score from the U-Test statistic in Equation 3.23 and Equation 3.24. These describe how the variability in SSST, as measured by the $F$ and $U$-Tests, depends on the variation in local physical parameters over spatial scales of <$1$ km and temporal scales of <$3$ hours. In this section a comparison is made between the above dependencies for the variability in SSST and the parameterisations for the bulk-skin temperature difference, $\Delta T$, described in Chapter 1, to investigate whether the two are governed by a similar set of factors. Such a conclusion would have important applications for understanding how SSST variability is related to $\Delta T$ and the ocean-atmosphere heat fluxes on the micro-scale level.

In Chapter 1 various existing parameterisations of the bulk-skin temperature difference, $\Delta T$, were outlined. This section will now examine each parameterisation to determine the dependency of $\Delta T$ on local physical parameters. This will facilitate a comparison of the dependency of SSST variability with the dependency of $\Delta T$.

The Saunders (1967) and Hasse (1971) parameterisations both feature the net heat flux, $Q$, across the atmosphere/ocean interface. This is given by:

$$Q = E + H + S + L$$

Equation 3.25

$S$ is the net incoming short-wave solar radiation, equivalent to the measured short-wave solar flux. $L$, the net long wave radiative heat loss is given by
\[ L = \epsilon_{SKY} \sigma T_{SKY}^4 - \epsilon_{SUR} \sigma T_{SUR}^4 \]

**Equation 3.26**

Where \( T_{SKY} \) is the sky temperature, \( T_{SUR} \) is the sea surface temperature and \( \epsilon_{SKY} \) and \( \epsilon_{SUR} \) are the spectrally integrated emissivity of the sky and sea surface respectively. Therefore, it can be written that \( L \) is a function of \( T_{SKY}^4 \) minus a function of \( T_{SUR}^4 \). That is:

\[ L = f(T_{SKY}) - f(T_{SUR}) \]

**Equation 3.27**

The latent heat flux, \( E \), is given by

\[ E = C_L \rho L_{EVAP} u (q_a - q_s) \]

**Equation 3.28**

Where \( C_L \) is the evaporation flux coefficient; \( \rho \) is the air density; \( u \) the mean wind velocity; \( q_a \) and \( q_s \) are the water vapour mixing ratios of the atmosphere and sea surface; \( L_{EVAP} \) the latent heat of evaporation.

\( L_{EVAP} \) and \( C_L \) are independent coefficients. The air density \( \rho \) can be written in terms of the air pressure \( P \) as:

\[ P = \rho g \]

**Equation 3.29**

Therefore,

\[ P = f(\rho) \]

**Equation 3.30**

The water vapour mixing ratio of the atmosphere \( q_a \), is a function of the relative humidity, RH:

\[ q_a = f(RH) \]

**Equation 3.31**
And the mean wind speed $u$ is a function of the wind speed, WS:

$$u = f(WS)$$  

**Equation 3.32**

Therefore Equation 3.28 can be re-written using Equation 3.29 to Equation 3.32:

$$E = f(P, WS, RH)$$  

**Equation 3.33**

The sensible heat flux, $H$, is given by

$$H = C_H \rho c_p u (\theta - T_{SUR})$$  

**Equation 3.34**

Where $C_H$ is the heat flux coefficient; $c_p$ is the specific heat of air at a constant pressure; $T_{SUR}$ is the sea surface temperature; $\theta$ is the overlying potential air temperature and is a function of the air temperature:

$$\theta = f(\Delta T)$$  

**Equation 3.35**

So Equation 3.34 becomes:

$$H = f(P,u,\Delta T,T_{SUR})$$  

**Equation 3.36**

Using Equation 3.27, Equation 3.33 and Equation 3.36 in Equation 1.3:

$$Q = f(R_{SKY}) - f(T_{SUR}) + f(P, WS, RH) + f(SF) + f(P, u, \Delta T, T_{SUR})$$  

**Equation 3.37**

The Saunders (1967) parameterisation of $\Delta T$ is given by
\[
\Delta T = \frac{-\lambda(v/\kappa)Q}{u^*c_p\rho}
\]

**Equation 3.38**

Where \( v \) is the kinematic viscosity; \( \kappa \) is the coefficient of molecular diffusion; \( Q \) is the net longwave, sensible and latent heat flow measured across the stagnant layer; \( u^* \) is the friction velocity of the upper ocean; \( c_p \) and \( \rho \) is the specific heat capacity and density of seawater respectively; \( \lambda \) is a wind dependent dimensionless coefficient. Therefore:

\[\lambda = f(WS)\]

**Equation 3.39**

Applying Equation 3.37 and Equation 3.39 to Equation 3.38:

\[
\Delta T = f(WS, TS_{sky}, TS_{sur}, P, RH, SF, AT)
\]

**Equation 3.40**

The second parameterisation of \( \Delta T \) is the Hasse (1971) model is given by:

\[
\Delta T = C_1 \frac{Q^*}{U} + C_2 \frac{Q_s}{U}
\]

**Equation 3.41**

Where \( C_1 \) and \( C_2 \) are derived coefficients; \( Q^* \) is the net heat flux through the surface (excepting solar flux); \( U \) is the mean wind speed at 4m above the sea surface; \( Q_s \) is the solar flux. Using Equation 3.37 the Hasse parameterisation becomes:

\[
\Delta T = f(WS, TS_{sky}, TS_{sur}, P, RH, SF, AT)
\]

**Equation 3.42**

This is the same as Equation 3.40. Thus, interestingly, is can be concluded that \( \Delta T \) is dependent on the same set of physical parameters, independent of which parameterisation is being considered. In order to compare this dependency with SSST variability it is necessary to re-write Equation 3.42 in terms of \( T_{sur} \), as this can be equated to the SSST variability:
\[ T_{\text{SUR}} = f(WS, T_{\text{SKY}}, \Delta T, P, RH, SF, AT) \]

Equation 3.43

The F-Test statistic and the Z score from the U-Test are both measures of the variability of \( T_{\text{SUR}} \) and so should have the same form of the above equation. However, it was found in Equation 3.23 and Equation 3.24 that they have the form:

\[ F = f(WS, SF, RH) \]

and

\[ Z = f(WS, SF, AT, RH, ST, \Delta T, t) \]

The Z score from the U-Test appears to have greater agreement with Equation 3.43 than the F-Test. The data from MuBEx ’95 did not show any significant dependency of Z on the variation in air pressure, which the parameterisations of \( \Delta T \) suggest should be the case. This may be due to lack of the range of air pressure encountered during any one transect. The dependency of Z on the time between TIC images, \( t \), is a probably a measure of the change in physical conditions between those images due to the RV Dai-Ni Misago moving in space over this time. Unfortunately, currents, wind variation and other factors that move the RV around, even with the engines off and the anchor/tethering deployed, mean that it is very difficult to decouple spatial differences between TIC images from the time between image.

The failure of the F-Test to produce a similar level of agreement with the \( \Delta T \) parameterisations suggests that something more significant than the limitations of the dataset is responsible. It is likely that the non-normality of the TIC image SSST data being tested, or the different size of some sets of TIC image pixel population have affected the F-Test values produced. This is despite the checks made on the results as described in 3.4.2.2. As a further test of the validity of the F-Test, the differences between U and F-Test results for the same TIC mage pairs are examined in the next section.

In conclusion, the variability of SSST at small scales, as quantified by the U-Test, exhibits the same dependency on local physical factors as parameterisations of the bulk-skin temperature difference, \( \Delta T \), and by extension, the ocean-atmosphere heat fluxes. This has implications, not only for understanding micro-scale ocean atmosphere processes, but for understanding how representative \textit{in situ} radiometric SSST measurements are for validating satellite-
measured SSST covering pixel areas of 1km² or more. The F-Test appears to produce differing results that do not show good agreement with the parameterisations of ΔT. The next section concentrates on a direct comparison of results from the F and U-Tests to investigate this disagreement in more detail.

### 3.4.7 Comparison of F and U-Tests results

The F and U-Tests have been defined in this chapter as a measure of SSST variability. Both the tests were applied to the TIC data in order to act as a check on the validity of each test. The previous section has found a deviation between results from the F-tests and those for the U-test. This sub-section further analyses this discrepancy by performing a direct comparison of the results from the two tests. In order to examine the co-incidence of the results of the U and F-Tests the equivalent confidence level for each test pair was computed. This was achieved by plotting all the F or Z values given for the appropriate degrees of freedom in the respective tables of the F and Z distributions (Lindman, 1974 and Hammond, 1978). A curve fitting routine applied to these plots yields an accurate confidence value for intermediate F and Z test values (Lindman, 1974).

The confidence levels determined by the F and U-Tests respectively differed from each other by a maximum of 10%. Although this demonstrates that there is a disagreement between the two tests for variability, it is concluded that there is a high degree of correlation between the F and U-Test results. Therefore, this analysis suggests that both tests are valid for use as a measure of variability for the TIC SSST data set.

Possible sources of error have been discussed in sub-sections 3.4.2.2 and 3.4.3.1. For the parametric F-Test, non-normality of the distribution of the test data, unequal variances and unequal sample sizes can cause errors. Application of the checks for these given in Lindman (1974) resulted in the conclusion that the F-Test was robust for the TIC SSST data being tested. For the non-parametric U-Test, the main source of error is in the determination of the accuracy of the TIC pixel values prior to application of the U-Test. Again, trials with differing levels of accuracy produced little variation in the U-Test results and in sub-section 3.4.3.1 it was concluded that it was legitimate to use an accuracy of ± 0.05K for the TIC data.

The 10% maximum difference between confidence levels determined by the F and U-Tests could be due to the sources of error summarised in the above paragraph. However, in sub-
section 3.4.6 a difference was found between the dependency of $F$ on physical parameters, as compared to that of the $Z$ score from the U-Test and that derived from existing parameterisations of $\Delta T$. This, coupled with difference between the $F$ and U-Test results described in this section leads to the conclusion that the F-Test is in error. This is probably due to the effect of non-normality and differing population sizes on the F-Test. However, the F-Test will still be applied to large-scale SSST and BSST data in Chapter 4 to confirm the supposition that produces these errors.

### 3.5 Variability of TIC image Brightness Temperature: implications for satellite SSST validation

#### 3.5.1 Methodology: investigation of standard deviation

The objective of this section is to find an optimum recommendation of observation time and area for *in situ* radiometric measurements to validate satellite SSST. The last section was able only to produce a dependency, rather than a parameterisation of SSST variability in terms of physical parameters. A more comprehensive *in situ* data set should in principle enable this parameterisation to be determined, allowing a measure of local SSST variability to be calculated during validation cruises. A confidence level could then be assigned as to whether the SSST measurement is typical of the satellite pixel being validated.

In the absence of a parameterisation of SSST variability, a different methodology for measuring SSST variability needs to be developed. The approach adopted for this analysis was to find the number of TIC images required to return a minimum variance from the mean SSST averaged over all those images. If the variance (and therefore variability) within a radiometric measurement can be minimised, the resulting mean SSST has a higher probability of being a “typical” SSST for that area of ocean.

The procedure found the variance of each TIC image in a transect. If there are $I$ images in a transect and the variance of each image is $\sigma_i^2, \sigma_2^2, \sigma_3^2, ..., \sigma_i^2$, with each image having $n_1, n_2, n_3, ..., n_I$ number of pixels, then the mean variance over all the images is:

$$
\sigma_i^2 = \frac{\sum_{i=1}^{I} \sigma_i^2}{I}
$$

Equation 3.44
A radiometer is used to obtain a validation measurement equivalent to the area of one TIC image with the same integration time as the TIC for the radiometric measurement. The derived SSST measurement would have a 95% probability of falling in the range \( \bar{x} - \sigma_1 \leq T \leq \bar{x} + \sigma_1 \), where \( \bar{x} \) is the mean of the standard deviations across each single TIC image, given in Equation 3.44, and \( \bar{x} \) is the mean SSST across all images.

Now consider the effect of lengthening the temporal and spatial coverage of the validation radiometer. The radiometer is pointed at the sea for twice as long as before, thereby doubling the area of the sea observed and obtaining a radiometric measurement equivalent to two TIC images. The SSST derived from this measurement would have a 95% probability of being in the range \( \bar{x} - \sigma_2 \leq T \leq \bar{x} + \sigma_2 \) where \( \sigma_2 \) is the mean standard deviation over all adjoining pairs of TIC images. Computationally, processing time can be saved by calculating the standard deviations of adjoining pairs of TIC images from their individual variances. These individual variances have already been obtained for use in Equation 3.44. The general form of the formula for obtaining a global variance from individual image variance is given below.

\[
\frac{\sum_{i=0}^{N} (n_i - 1) \sigma_i^2}{N - 1}
\]

**Equation 3.45**

Where \( N = \sum_{i=0}^{I} n_i \) and \( I \) is the number of images being combined. So to calculate the variance of an adjoining pair of TIC images, Equation 3.45 becomes:

\[
\frac{(n_1 - 1) \sigma_1^2 + (n_2 - 1) \sigma_2^2}{N - 2}
\]

**Equation 3.46**

If we continue to scale up the number of TIC images that the "validation radiometer" is sampling, a plot can be obtained of the number of TIC images (which translates into sample time) against mean standard deviation. Where the mean standard deviation is minimised, a "validation radiometer" obtaining a measurement or series of measurements over the
corresponding sampling time/ocean surface area has the highest probability of obtaining a “typical” SSST for the area, in comparison to the other sampling periods on the plot.

To obtain these plots, the formula for obtaining a global variance from the variance of individual images (Equation 3.45) was also used to obtain global variances for adjoining sets of 3 TIC images, then 4 and so on. The resulting plots for 20 MuBEx 95 transects of varying duration are shown in Figure 3.19 to Figure 3.21. A greater number of transects was processed to produce these results than for the $F$ and U-Tests. This is because Brightness Temperature data can be used for the investigation of standard deviation as the correction for downwelling longwave radiance will alter the individual pixel temperatures by the same value. Therefore, the correction for downwelling longwave radiance affects the TIC image mean and not the standard deviation. Use of SSST was necessary for the $F$ and U-Tests as the means of the images being tested are a factor in these tests for dissimilarity of images (Lindman, 1974 and Hammond, 1978).
Figure 3.19: Variance of SSST against sampling interval for transects from 26/7/95 to 2/8/95 from MuBEx 95
Figure 3.20: Variance of SSST against sampling interval for transects from 3/8/95 to 12/8/95 from MuBEx 95
Figure 3.21: Variance of SSST against sampling interval for transects from 12/8/95 to 23/8/95 from MuBEx 95

3.5.2 Results Analysis

The plots of standard deviation against sampling interval (Figure 3.19 to Figure 3.21) can be divided into four categories of behaviour. These are described in Table 3.12. An analysis was performed to ascertain whether there is any correlation between the trend in standard deviation and the measured physical parameters described in Table 2.2. As the behaviour of the standard deviation is measured over entire transects, it is problematic to know which aspects of co-incident meteorological data can be examined for correlation. The mean and standard deviation of each meteorological parameter for each transect was calculated to see if any behaviour categories were associated with high or low means or standard deviations of these parameters. However, no category produced any significant trend in this context. More qualitative analyses examining log books from the transects for cloud cover and sea state observations were undertaken, but again no correlation was found between these and the behaviour categorisation. A check on whether diurnal effects could be a factor resulted in no correlation between data collected at night, for example, and the behaviour of the standard deviation.
Table 3.12: Categorisation of trend for standard deviation (σ) against sampling interval for MuBEx 95 Brightness Temperature data

<table>
<thead>
<tr>
<th>Category Designation</th>
<th>Description</th>
<th>Transects exhibiting behaviour</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Indeterminate (due to brevity of transect)</td>
<td>26/07 1312, 02/08 0936, 11/08 2303, 12/08 0300</td>
</tr>
<tr>
<td>B</td>
<td>Continuous trend of increasing σ with sampling interval</td>
<td>27/07 0940, 29/07 0945, 30/07 0618, 03/08 1115</td>
</tr>
<tr>
<td></td>
<td></td>
<td>08/08 0922, 12/08 0054</td>
</tr>
<tr>
<td>C</td>
<td>Initial trend of increasing σ with sampling interval; for longer sampling</td>
<td>26/07 1006, 26/07 1201, 02/08 1042, 03/08 1222</td>
</tr>
<tr>
<td></td>
<td>intervals decreasing σ with increasing sampling interval</td>
<td>11/08 2043, 12/08 0611, 12/08 1938, 22/08 0618,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>23/08 1005</td>
</tr>
<tr>
<td>D</td>
<td>Initial trend of decreasing σ with increasing sampling interval; for longer</td>
<td>12/08 2203</td>
</tr>
<tr>
<td></td>
<td>sampling intervals, increasing σ with sampling interval</td>
<td></td>
</tr>
</tbody>
</table>

In conclusion, 3 types of trend of standard deviation against sampling interval have been identified. One trend (B) suggests that standard deviation of Brightness temperature is minimised over a short sampling period for radiometric measurements of the sea surface. Quantitatively this equates to a 10-second measurement of the sea surface. If a TIC image has an area of between 3.10m² and 3.94m² (see 3.3) and the TIC is sampling an image every 5 seconds, then this sampling interval represents a 6.20m² to 7.88m² area of the sea surface. Therefore, a “typical” SSST measurement for validating satellite measurements would be more likely to be obtained over short sampling times.

The second trend type (C) shows standard deviation increases to a peak and then declines with increasing sampling time. This would indicate that longer sampling periods of greater than 1500 seconds would have a higher probability of returning a “typical” SSST. This represents a sampled area of the ocean surface of between 930m² and 1182m². The third trend type (D) was only found in 1 of the transects of data. This trend is characterised by a minimum standard deviation at a sample period of 400 seconds, equivalent to an area of the ocean surface of 248m² to 315.2m².
No significant correlation between meteorological data or local sea and sky conditions can be found. However the data set is limited in terms of the range of meteorological conditions encountered and the temporal extent of observations. It should also be noted that errors may be present, especially as the mean of the standard deviations vary by the order of $10^{-4}$ K for the different integration times whereas the brightness temperatures from which the standard deviations are derived are accurate to ±0.1K. The very large number of the measurements from which the standard deviations are derived will reduce the likelihood of error but the results should still be treated with caution. Further investigations using radiometric imaging technology and co-incident meteorological observations to obtain a more extensive set of source data are suggested.

### 3.6 Conclusions

In this Chapter the variability of the skin sea surface temperature over small (less than 1km) scales has been investigated using a thermal infrared camera. The variability of SSST has been measured using the analysis of variance F-Test and the non-parametric Mann-Whitney U-Test to compare TIC images of SSST. These have both found that a high degree of variability of SSST exists over small spatial scales (<1km) and temporal scales (<3hours). Typically over 99% of image pairs tested over scales were found to disprove the null hypothesis that TIC image pairs are from the same population and that variations in SSST between two images can be accounted for by random fluctuation. Both tests have been found to the same confidence probability for testing the null hypothesis within a 10% error.

The values of the $F$ and U-Tests have been correlated against the variation in coincident meteorological parameters between TIC images. It has been found a dependency exists of $F$ on the variation between images in wind speed, solar flux and relative humidity and that the dependency has the functional form:

$$F = f(WS, SF, RH)$$

Equation 3.47
The relationship between the Z score, derived from the U-Test statistic, and the variation in wind speed, solar flux, air temperature, sky temperature, relative humidity was found to have the functional form:

\[ U = f(WS, SF, AT, RH, ST, \Delta T, t) \]

**Equation 3.48**

Existing parameterisations of \( \Delta T \) produce the dependency:

\[ T_{SUR} = f(WS, T_{SKY}, \Delta T, P, RH, SF, AT) \]

**Equation 3.49**

Therefore, the small-scale variability of SSST, as quantified by the U-Test, and the parameterisations of \( \Delta T \) are in very good agreement as to the dependencies of these factors on physical parameters. Furthermore, as the parameterisations of \( \Delta T \) include the ocean-atmosphere heat fluxes, then it is concluded that SSST variability is dependent upon these fluxes. The absence of an air pressure term in the U-Test dependency equation is attributed to the lack of variation in this parameter during MuBEx '95. The presence of time in the same equation is probably due to this being a proxy for variation in the other physical parameters due to the movement of the RV *Dai-Ni Misago*.

The F-Test appears to depend on fewer physical parameters than the U-Test. This, coupled with a comparison of the \( F \) and U-Test results for identical TIC image pairs have generated the conclusion that the F-Test is in error. It is postulated that the error is due to the non-normality of the TIC image SSST data being tested, or the different size of some sets of TIC image pixel population.

A further analysis of the TIC data has been carried out to ascertain the spatial and temporal scale of radiometric measurements which have the greatest probability of yielding a "typical" SSST for the transect of data. In order to achieve this, the standard deviations of SSST in the TIC images were investigated. The spatial/temporal scales of combinations of TIC images that produce a minimum value for the standard deviation were found. By minimising the standard deviation, radiometric measurements of SSST over these scales are more likely to produce a "typical" SSST.
Three types of trend of standard deviation against sampling interval were identified, producing minimum values for the standard deviation over sampling times of <10s 400s and >1500s respectively. These are equivalent to sampling areas of the sea surface of between 6.20m² and 7.88m², 248m² to 315.2m² and 930m² and 1182m². No significant correlation between these types of trend and meteorological data or local sea and sky conditions could be found. However this may be due to limited nature of the data set and further investigation using a more comprehensive in situ data set is recommended.

The recognition of the dependency of SSST variability on the same physical factors that govern AT and the ocean-atmosphere heat fluxes has many implications. It provides the foundation for using a more comprehensive in situ data set to parameterise SSST variability, as defined by the Mann-Whitney U-Test. This will enable small-scale SSST variability to be measured in directly and provide satellite SSST validation experts with a method of quantifying how typical their in situ measurements are of the satellite pixel SSST. Another implication of this work is in the understanding of ocean-atmosphere interaction on small scales. Furthermore, by quantifying SSST variability, a comparison of this variability at small and large-scales can be made. If the variability of SSST differs markedly at these diverse scales then climate modellers may need to assess whether any modifications are needed to reflect small-scale SSST variability, as most models utilise large-scale datasets.

The MuBEx 95 data set is limited in terms of the range of meteorological conditions encountered, the temporal extent of observations and the inshore nature of the experimental area. In the next Chapter, the F-Test and Mann-Whitney U-Test for variability of SSST are applied to longer-term mesoscale data. The results obtained so far for small spatial scale SSST observations are then compared with those for mesoscale SSST.
4. COMPARISON OF LARGE-SCALE TROPICAL PACIFIC SSST AND BSST DATASETS

4.1 Overview
This Chapter applies the $F$ and $U$-Tests to large-scale SST data that is used in climate modelling. In Chapter 5, an ocean model is used to test the effect of using a sea surface temperature dataset, which includes satellite-measured SSST, as opposed to purely BSST data. In this Chapter the data being tested by the $F$ and $U$-Tests is the same data which is applied in the climate model in Chapter 5.

The tests are applied to two sets of SST data. The first set of $F$ and $U$-Tests is conducted on the SSST data. The aim of this procedure is to compare the variability of large-scale SSST with that of small-scale SSST, which was investigated in Chapter 3. The second set of $F$ and $U$-Tests is applied to large-scale BSST and SSST data in order to test the hypothesis that, statistically, the two sets of data are from the same population. This would produce the hypothesis that the use of SSST instead of BSST data in climate modelling should have no significant effect.

The large-scale SST data used in this Chapter is drawn from two sources. The SSST data is from the Along Track Scanning Radiometer-1 (ATSR-1) instrument carried aboard the ERS-1 satellite. The BSST data is from the Combined Ocean Atmosphere Dataset (COADS). These data sets are described in the following sections.

4.2 The Along Track Scanning Radiometer-1 (ATSR-1) instrument
The Along Track Scanning Radiometer-1 (ATSR-1) instrument is an imaging radiometer carried aboard the ERS-1 (European Remote Sensing) satellite. The ERS-1 platform is in a near-circular, retrograde, sun-synchronous polar orbit at a mean height of approximately 780 km, producing an orbital period of about 100 minutes. The repeat period for overpasses is generally 3 days, although this has been altered to 35 and 168 days for some parts of the ERS-1 and ERS-2 missions. ERS-1 was launched in July 1991. Data collection was concluded by ESA in June 1996, two years in excess of the satellite's design lifetime and the satellite is now in hibernation mode. A successor radiometer, ATSR-2, became operational in 1995 while a
third, AATSR, will be launched in the year 2001. The analysis in this work uses data from ATSR-1.

The ATSR-1 instrument was designed for exceptional sensitivity and stability of calibration to enable the retrieval of an accurate global SST data set. This was achieved through use of low-noise infrared detectors cooled to their optimum operating temperature of less than 95 K by a Stirling cycle mechanical cooler. Detector noise is thereby minimised and single channel equivalent temperatures can be determined correct to ±0.05K.

As with the AVHRR (Advanced Very High Resolution Radiometer) instruments (Lauritson et al., 1979), a multi-channel approach to SST retrieval was used. The “along-track scanning” technique is used to provide two (nadir and forward) views of the surface enabling radiance from the surface to be measured through two different atmospheric path lengths. This allows an improved correction for atmospheric effects to be introduced into the SST retrieval algorithm. The ATSR-1 has three thermal infra-red channels (centred at 3.7μm, 10.8μm and 12μm), chosen to match those of the existing AVHRR series of instruments. In addition, the radiometer has a reflected infrared channel at 1.6μm in order to detect clouds by day. The 1.6 and 3.7mm channels employ photovoltaic indium antimonide (InSb) detectors and the 10.8 and 12.0mm channels use photoconductive cadmium mercury teluride detectors (PC CMT). These spectral channels are summarised in Table 4.1.

Table 4.1: ATSR-1 Spectral Channels

<table>
<thead>
<tr>
<th>Feature</th>
<th>Wavelength</th>
<th>Bandwidth</th>
<th>Detector type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cloud Clearing</td>
<td>1.6μm</td>
<td>0.3μm</td>
<td>PV InSb</td>
</tr>
<tr>
<td>SST retrieval</td>
<td>3.7μm</td>
<td>0.3μm</td>
<td>PV InSb</td>
</tr>
<tr>
<td>SST retrieval</td>
<td>10.8μm</td>
<td>1.0μm</td>
<td>PC CMT</td>
</tr>
<tr>
<td>SST retrieval</td>
<td>12.0μm</td>
<td>1.0μm</td>
<td>PC CMT</td>
</tr>
</tbody>
</table>

A further feature of the design of ATSR-1 to ensure a stable calibration was the use of two on-board stable, high-accuracy blackbody calibration targets to enable continuous calibration of the infrared channels of the instrument. This has produced calibration of the radiometer to less than 0.1K (Mason et al. 1996). The associated calibration algorithms are described later in this Chapter. A detailed description of the data from ATSR-1 can be found in Mutlow et al., 1999. The nominal instantaneous field of view (IFOV) pixel size is 1 km² at the centre of
the nadir swath and 1.5 km × 2 km at the centre of the forward swath. A 75 ms signal integration period produces each pixel.

The primary reasons for utilising ATSR-1 data in this study are that the radiometer is designed to measure very accurate SST and that the retrieval algorithm is tuned to SSST, unlike many earlier satellite radiometers. The calibration scheme is one of the key components of ensuring the accuracy of ATSR-1 data. The calibration of ATSR-1 involves determining the linear relationship between the radiance and detector counts from each channel. The signal count from a radiometer channel observing a blackbody target at temperature $T_{bb}$ is

$$S(T_{bb}) = GL(T_{bb}) + S_0$$

\[ \text{Equation 4.1} \]

where $G$ is the radiometric gain, $L(T_{bb})$ is the radiance from a target (i.e. the Planck function integrated over the filter passband), and $S_0$ is the radiometric offset of the channel. Earlier satellite radiometers, such as AVHRR, allow the instrument to view a zero radiance target, such as a cold space view, to determine the radiometric offset $S_0$. The radiometer then views a hot calibration target to determine the gain of the channel. The gain of the system is given by

$$G = \frac{S_{cold} - S_0}{L_{cold}}$$

\[ \text{Equation 4.2} \]

However, radiometers generally have some non-linearity, which may also change over time. To minimise errors from non-linearity for ATSR-1 two approaches were adopted. The first was to carefully design the signal processing electronics and undertake pre-flight determination of the non-linearity for "beginning-of-life" and "end-of-life" conditions on the satellite. The second (and principal) approach was to use two on board calibration targets maintained at temperatures above and below the range of expected SSTs (Mutlow et al, 1999). This ensures that the calibration is most precise over the normal range of SST. As
linearity is assumed over a small span of measurement range errors from non-linearity in the system are minimised.

Retrieval of SSST from ATSR-1 data is performed using the Rutherford Appleton Laboratory's data-processing scheme SADIST (Synthesis of ATSR Data Into Sea-surface Temperatures), (Závody et al., 1994). This scheme enables the infrared brightness temperatures measured by ATSR-1 to be converted into the best estimate of the SSST for each cloud-free pixel over the sea and thus compiled into a SSST image at 1km resolution. The scheme is continuously being validated and refined.

Whenever possible, both nadir and forward view data are used in the retrieval process. There is also provision for use of nadir-only brightness temperatures, as the forward view is more likely to suffer cloud contamination due to the larger sampling area in forward view pixels. The presence of a 3.7μm channel on ATSR-1 allows a 3 channel retrieval scheme to be applied in some cases. However during daytime this channel is substantially contaminated by reflected solar radiation, so a 2 channel retrieval scheme is used. Furthermore, on May 27th, 1992 the 3.7μm channel of ATSR-1 failed. Therefore, SSST retrievals from this date onwards utilise only the 11 and 12μm channels (Murray et al. 1998).

The ATSR-1 data used in this study is subject to several potential sources of error. The cloud detection algorithm used to pre-process the data has undergone various subsequent improvements and the data therefore may include errors from cloud contamination. The SSST retrieval model may also contain errors, particularly with respect to allowing for the aerosol content in the atmosphere, which increased significantly during the data-gathering period for ATSR-1 due to the volcanic eruption of Mount Pinatubo in 1991.

4.3 The ATSR-1 Spatially-Averaged Sea-Surface Temperature Product

The Averaged Sea-Surface Temperature (ASST) product contains spatially averaged sea-surface temperatures, at 0.5° resolution, using both nadir-only and dual-view retrieval algorithms. The ASST data is generated using the SADIST v.500 software (Bailey, 1993) and is derived from collocated pixel data.
The ASSTs are derived via several steps: 10-arcminute brightness temperatures are compiled by averaging 1km x 1km brightness temperatures. The retrieval algorithms (Zavody et al. 1994, 1995) are then used to derive SSTs. The 0.5° ASST product is then computed by averaging these 10-arcminute SSTs, with nine of the 10-arcminute cells being contained in a 0.5° cell. Three products are generated: nadir-view only values; dual-view only values; and SST values using a dual-view retrieval if possible, and nadir-only retrieval otherwise.

In this study, the latter product is utilised. This maximises coverage while accepting that values derived using nadir-only data may introduce larger errors. The dual-view only product represents the most accurate retrieval but typically will have 15% less coverage than when nadir-only values are included (Murray, 1995). At the time of processing, ATSR-1 ASST values were available for the 4 years from August 1991 to July 1995.

4.4 The Comprehensive Ocean-Atmosphere Data Set (COADS)

The large-scale BSST data used both for the SST variability analysis in this Chapter, and in Chapter 5 to drive an ocean model, are extracted from the Comprehensive Ocean-Atmosphere Data Set (COADS). This is a combined database of ship and buoy BSST and meteorological observations, dating from 1854. The project is a co-operative effort between the National Oceanic and Atmospheric Administration (NOAA) and the National Center for Atmospheric Research (NCAR).

The data used here are from the extended COADS 1a product (Woodruff et al. 1993, Woodruff et al. 1998). This consists of a monthly mean BSST values for 2° latitude x 2° longitude boxes calculated using observed data falling within each box. Other basic observed variables in COADS include air temperature, wind velocity, humidity (wet bulb or dew point temperature), barometric pressure, cloudiness, weather, and wave and swell fields.

Throughout the original COADS Release 1 individual observations are quality controlled by "trimming". This is a process by which individual observations are tested against upper and lower quality control limits defined for each 2° grid square and month. For Release 1, the trimming limits are set at the 3.5 standard-deviation (sigma) level using three climatological periods (1854-1909; 1910-49; 1950-79).
The 3.5 sigma trimming limits have been shown to produce problems when attempting to detect or study extreme climate anomalies such as the 1877-1878 and 1982-83 El Niño/Southern Oscillation (ENSO) events (Wolter et al., 1989; Wolter, 1997). Furthermore, changes in spatial and temporal sampling, instrumentation and observational practices, and data processing can result in inhomogeneities within the ship-sourced data. The validity of combining data from ships with that from buoys, sea temperatures from oceanographic profiles, and data from fishing vessels has also been questioned, particularly with respect to wind and temperature data (Barnett, 1984; Wilkerson and Earle, 1990; Woodruff et al., 1993).

In an attempt to begin addressing these concerns, the Release 1a COADS product is available as a "standard" or "enhanced" version. The standard version applies the 3.5 sigma trimming limits, but the data is restricted mostly to ship observations. The "enhanced" version includes automated platform data in addition to ships, and is processed using a 4.5 sigma trimming limit. This increases spatial coverage and provides a more comprehensive representation of extreme climate anomalies.

The ocean model in Chapter 5 covers the equatorial Pacific Ocean. This region exhibits anomalies in SST during ENSO events. Given the aforementioned choice of data sets, the enhanced version of COADS Release 1a is used in this Chapter and Chapter 5 in order to maximise the probability that the BSST dataset will represent ENSO climate anomalies.

4.5 Pre-processing

It was stated earlier in this Chapter that the large-scale SSST and BSST data being tested for variability are the same data that will be used to force the ocean model in Chapter 5. These data must therefore be compatible with the data requirements of the model. A 15-year SST climatology, running from 1981 to 1995, is used in the model (Rathman, 1998). This SST covers a $114^\circ \times 42^\circ$ area of the equatorial Pacific Ocean on a grid at 1.5° resolution. The longitude limits of this area run from 121.5°E to 67.5°W and the latitude limits are from 31.5°N to 31.5°S.

The individual monthly mean SST measurements used to calculate the climatology are not always available. ATSR-1 ASST data may be missing due to instrument maintenance, orbital manoeuvres or cloud contamination. COADS data may not be available due to lack of observations. However, the model requires a full set of data to function.
Two phases of pre-processing were implemented. The first phase uses a weighted nearest neighbour routine to fill in missing data gaps in the unprocessed ASST and COADS products. The data then needs to be scaled to the required 1.5° grid. For the 0.5° grid resolution ASST data this is a direct averaging calculation. For the 2° grid resolution COADS SST, a bilinear interpolation routine was applied using the IDL programming language. These processes result in two data sets: the ATSR values are given as a 114 x 42 cell grid of monthly mean SSSTs for each month from August 1991 to July 1995; the COADS data is a 114 x 42 cell grid of monthly mean SSSTs for each month from January 1981 to December 1995.

4.6 Methodology

Given the two sets of data produced by the pre-processing described in 4.5, several comparisons were made using the $F$ and U-Tests for variability. Following from the work described in Chapter 3, the variability of the SSST data was tested. Each 114 x 42 set of monthly mean SSSTs was treated as being analogous to the TIC images in the small-scale SSST study. Applying the same methodology utilised in Chapter 3, each 114 x 42 large-scale "image" of SSST was tested against every other SSST "image". Both the $F$ and U-Tests were used in this process and the results are described in 4.7 and 4.8.

The next set of variability analysis uses the SSST data from ATSR-1 to test the hypothesis that monthly means for the same month in different years will be from the population. For the 1.5° 4 year 114 x 42 cell SSST data set each "image" from January in one year was tested against other "images" from January in the other years. This procedure was then repeated for February, March and so on. As there are 4 years of data available for each month, the number of tests performed (excluding identities) for each month is 6. Therefore a total of $12 \times 6 = 72$ tests were performed in this analysis.

The final set of tests concentrated on examining whether the large-scale BSST and SSST data, which are used to force the model in Chapter 5, are from the same population. If the $F$ or U-Tests show that the monthly mean BSST and SSST data for each month are from the same population then we might expect the model to perform in a similar manner when driven by climatologies derived from BSST or SSST data. The analysis performed consisted of applying the $F$ and U-Tests to pairs of BSST and SSST monthly means for each of the 48 months where SSST data was available.
4.7 F-Test Results

4.7.1 All combinations of SSST pairs

A total of 1128 F-Tests was performed, representing all the combinations of 2 SSST data set "images" possible from an overall set of 48 "images". This figure does not include identities or reciprocal tests, the latter resulting in the same $F$ value as the "forward" comparison of SSST "images" pairs. This calculation for the number of test combinations possible is statistically expressed as $48C_2$. As with the F-Tests in Chapter 3, the check for the effect of unequal variances (section 3.4.2.2) was applied to all SSST test pairs. As all test population sizes are equal (there is no missing data), the check for the effect of unequal sample sizes was not necessary. The test for the effect of unequal variance produced a maximum error probability of 4.2%.

The mean, minimum and maximum and standard deviation of the resulting $F$ values is shown in Table 4.2. It will be recalled from Chapter 3 that the $F$ value for 99.5% confidence of the two test populations being dissimilar is $F = 7.879$. The $F$ values that fall below this threshold represent test populations that cannot be said to be dissimilar with a 99.5% level of confidence. As Table 4.2 shows, no F-Test value exceeded 0.104. This is equivalent to a 35.54% confidence level of dissimilarity. That is, the two test populations are likely to be from the same global population. The mean $F$ value in Table 4.2 of $F = 0.013$, and the minimum value of $F = 2.06 \times 10^{-8}$ represent far lower confidence levels of dissimilarity. This indicates that the large-scale SSST data from ATSR-1 is very homogeneous, particularly in comparison with the high variability of the small-scale SSST data tested in Chapter 3.

<table>
<thead>
<tr>
<th>F-Test Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
</tr>
<tr>
<td>Min</td>
</tr>
<tr>
<td>Max</td>
</tr>
<tr>
<td>SD</td>
</tr>
</tbody>
</table>

4.7.2 Comparison of SSSTs for particular months

The four-year monthly SSST data set enables six unique pairs of large-scale "images" of SSST monthly means to be subjected to the F-Test for each month. This is expressed
statistically as $^4C_3$. As there are 12 months in a year, the number of tests performed in this analysis totalled 72.

The mean, minimum and maximum and standard deviation of the resulting $F$ values is shown in Table 4.3. The maximum $F$ values of $F = 0.0234$, and the mean $F$ value of $F = 0.0048$, are even lower than those for all SSST combinations evaluated in section 4.7.1. This indicates that F-Tests comparing large-scale SSST for each particular month over the years covering 1991 to 1995 result in a probability that these data exhibit even less variability than that for all SSST pairs over the same period. This result might be expected, as a comparison of two sets of monthly mean SSSTs for January, each one from a different year, should result in the two test populations being more similar than, say, a comparison of the monthly mean SSSTs from January and June.

Table 4.3: F-Test results summary for comparisons of SSST pairs for particular months

<table>
<thead>
<tr>
<th>F-Test Value</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.0048</td>
</tr>
<tr>
<td>Min</td>
<td>2.16x10^-6</td>
</tr>
<tr>
<td>Max</td>
<td>0.0234</td>
</tr>
<tr>
<td>SD</td>
<td>0.0058</td>
</tr>
</tbody>
</table>

An argument against this explanation is that the data being tested is from the equatorial Pacific Ocean and should therefore not be as dependent on the annual seasonal climate cycle experienced in higher latitudes. However as the area covered by the data extends between 31.5°N and 31.5°S there is some seasonal pattern to SST in this area. To further investigate the variability of SSST for specific months, the mean $F$ value for the comparisons of SSST "images" in each month was found. These means are plotted in Figure 4.1. This suggests that the summer months of May to September exhibit the lowest variability. However, any conclusions drawn from this analysis must contain the qualifier that the means are based on a sample size of only 6.
4.7.3 Comparison of BSST and SSST data for each month

The final set of F-Tests are a set of 48 direct comparisons between the SSST and BSST monthly means for each month in the 4 years of available SSST data. The summary of these tests is shown in Table 4.4. The high mean, maximum and minimum $F$ values for this set of tests far exceeds the threshold of $F = 7.879$ which equates to a 99.5% confidence level of two test populations being dissimilar. The F-Test therefore suggests that all the month-by-month comparisons of monthly BSST means with monthly SSST means result in the conclusion that both sets of data are dissimilar and do not come from the same global population.

Table 4.4: Summary of F-Test results for direct comparisons of BSST and SSST

<table>
<thead>
<tr>
<th>F-Test Value</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>24353</td>
</tr>
<tr>
<td>Min</td>
<td>23757</td>
</tr>
<tr>
<td>Max</td>
<td>24809</td>
</tr>
<tr>
<td>SD</td>
<td>255</td>
</tr>
</tbody>
</table>
4.8 U-Test Results

4.8.1 All combinations of SSST pairs

As with the F-Tests, a total of 1128 U-Tests was performed. The mean, minimum and maximum and standard deviation of the resulting Z values is shown in Table 4.5. The Z score for a 99.5% confidence level of the two test populations being dissimilar is $Z = 2.8$. Z scores which are lower than this threshold represent test populations that cannot be said to be dissimilar with a 99.5% level of confidence.

Table 4.5 shows that the mean Z score was 38.15, which is far in excess of the $P(.995)$ threshold of 2.8. This therefore suggests that the null hypothesis is rejected, and most of the SSST test pairs are dissimilar. As the standard deviation from the mean is 27.25, 68% of the Z scores lie in the range $10.9 \leq Z \leq 65.4$, assuming a normal distribution about the mean of the Z scores. This further supports the interpretation that a large majority of SSST test populations are dissimilar.

However, the minimum Z score is 0.09, equivalent to a confidence value of 36.8% that the populations are dissimilar. So some test pairs do exhibit similarity and lower variability, although these are clearly a minority of large-scale monthly mean SSST combinations. In fact, only 5.1% of SSST test pairs produce Z scores lower than the $P(.995)$ value of 2.8. This is a higher proportion than the 0.907% of U-Tests applied to the small-scale SSST data in Chapter 3 (3.4.3.2). This indicates that while the large-scale monthly mean SSST data from ATSR-1 has a high degree of variability from month to month, the small-scale SSST data tested in Chapter 3 has an even greater variability.

Table 4.5: U-Test results summary for all combinations of SSST pairs

<table>
<thead>
<tr>
<th>Z Score Values</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>38.15</td>
</tr>
<tr>
<td>Min</td>
<td>0.09</td>
</tr>
<tr>
<td>Max</td>
<td>111.11</td>
</tr>
<tr>
<td>SD</td>
<td>27.25</td>
</tr>
</tbody>
</table>
4.8.2 Comparison of SSSTs for particular months

As with the F-Test (section 4.7.2) U-Tests were performed on 72 combinations of SSST data in this analysis. The results (Table 4.6) again show that the majority of Z scores exceed the $\text{P}(.995)$ Z score of 2.8. In this analysis, the values of the mean and standard deviation show that 68% of the Z score lie in the range $7.63 \leq Z \leq 50.59$. While this implies a high degree of variability between monthly mean SSSTs for the same month in different years, the mean Z score for this analysis is lower than that for all SSST combinations (Table 4.5). Furthermore, the proportion of Z scores exceeding the value for $\text{P}(.995)$ of $Z=2.8$ is lower than that found for all SSST combinations in 4.8.1. A total of 88.9% of U-Tests reject the null hypothesis that the test pairs are statistically similar. This compares to a 94.9% proportion for all SSST combinations.

These findings support those of section 4.7.2 that comparison of large-scale monthly mean SSST data for the same month but from different years covering the period 1991 to 1995, result in a probability that these data exhibit less variability than that for all SSST pairs over the same period.
Table 4.6: U-Test results summary for comparisons of SSST pairs for particular months

<table>
<thead>
<tr>
<th></th>
<th>Z Score Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>29.11</td>
</tr>
<tr>
<td>Min</td>
<td>0.09</td>
</tr>
<tr>
<td>Max</td>
<td>79.11</td>
</tr>
<tr>
<td>SD</td>
<td>21.48</td>
</tr>
</tbody>
</table>

As in section 4.7.2, to further investigate the variability of SSST for specific months, the mean $F$ value for the comparisons of SSST "images" in each month was found. These means are plotted in Figure 4.2. This suggests that the months of August to October exhibit the lowest variability. Again, any conclusions drawn from this analysis must contain the qualifier that the means are based on a sample size of only 6.

![Figure 4.2: Z score means for SSST comparisons for individual months](image)

4.8.3 Comparison of BSST and SSST data for each month

This analysis compares each of the 48 sets of SSST monthly means directly with the corresponding BSST monthly means. A summary of the resulting Z score values is shown in Table 4.7. As the maximum and minimum values of the Z score show, all U-Tests in this value are far in excess of the value for $P(.995)$ of $Z=2.8$. Thus, all the U-Tests reject the null hypothesis that the SSST and BSST monthly means are similar. That is, the BSST and SSST monthly means for equatorial Pacific Ocean are dissimilar.
Table 4.7: Summary of U-Test results for direct comparisons of BSST and SSST

<table>
<thead>
<tr>
<th></th>
<th>Z Score Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>316.4</td>
</tr>
<tr>
<td>Min</td>
<td>160.0</td>
</tr>
<tr>
<td>Max</td>
<td>488.0</td>
</tr>
<tr>
<td>SD</td>
<td>74.0</td>
</tr>
</tbody>
</table>

4.9 Comparison of F and U-Test Results

The comparison of all possible SSST monthly mean pairs produced considerably different results for the F-Test and U-Tests (sections 4.7.1 and 4.8.1). The F-Test analysis found that all the test pairs were likely to be from the same global population, using 99.5% confidence $F$ value as the test threshold for rejecting the null hypothesis. The U-Test results found that only 5.1% of SSST test pairs produced Z scores lower than the 99.5% confidence Z value. Therefore, the $F$ and U-Tests were in agreement for just 5.1% of test pairs.

The F-Test found that all SSST test pairs for the same month in different years produced $F$ values below the 99.5% threshold for rejecting the null hypothesis. So, the SSST test pairs are likely to be similar. These tests are a sub-set of the tests for all SSST combinations, so it should be no surprise that no F-Test exceeded the P(.995) threshold. Again, the U-Test results disagree with those of the F-Test with 11.1% of SSST test pairs agreeing with the F-Tests and producing Z scores below the 99.5% confidence threshold. However, both $F$ and U-Tests produced lower $F$ and Z scores than those for all SSST combinations, indicating greater homogeneity in monthly mean SSST from the same months in different years.

Examining the mean $F$ and U-Test values for individual months (re-plotted together in Figure 4.3 to facilitate comparison) it can be seen that there is agreement in the trend of mean values for months from October to April. The tests also agree that the minimum $F$ and Z scores occur in August and September, although there is some discrepancy in the trend in May, June and July.
Comparing corresponding SSST and BSST monthly means, the $F$ and U-Tests agree that the null hypothesis can be rejected with a confidence level in excess of 99.5% for all test pairs, indicating a high level of inhomogeneity between the BSST and SSST data sets.

The disparity between the results of the $F$ and U-Tests both in this Chapter and Chapter 3 is most likely to be the result of non-normality or unequal variances between the test populations producing an error in the F-Test. This is despite the checks for these possible problems described in Lindman (1974) being applied. As the U-Test is a distribution-free process, it is not affected by these factors and is therefore more reliable than the F-Test, despite having 95% of the power of the F-Test.
4.10 The dependency of the F and U-Test values for SSST data on physical parameters

In Chapter 3, the dependency of the $F$ and $U$-Test values on measured parameters such as wind speed, humidity, air temperature, solar flux and so on, was investigated and the nature of the dependency empirically derived. Similar data is available as part of the COADS dataset as a monthly mean value for each grid square. However, the values of the physical parameters for the small-scale analysis in Chapter 3 were those across an entire TIC image. If the analogy were to be extended to the large-scale analysis performed in this Chapter, then a mean value for these parameters across the entire Tropical Pacific Ocean would be required. It is believed that this would not be a valid exercise. However, a large-scale analysis using a more moderate scale, such as $100m \times 100m$ grid size SSST measured by an aircraft would be more likely to enable a valid comparison.

4.11 Conclusions

This Chapter has examined the variability of large-scale SSST and BSST monthly means with two main objectives. The first is to see if large-scale SSST exhibits the same behaviour as the small-scale SSST data examined in Chapter 3. The second is to investigate whether large-scale SSST monthly means exhibit a high degree of variability in comparison to BSST monthly means.

The tests found that the large-scale ATSR-1 SSST monthly means for the Tropical Pacific between August 1991 and July 1995 are more homogenous than small-scale SSST. The F-Test results show that the large-scale SSST has a very low degree of variability, whereas the U-Test results indicate a high degree of variability, but lower than that for small-scale SSST. This has the important conclusion for climate researchers that any large-scale monthly-averaged SSST used for climate modelling does not reflect the variability of small-scale SSST. This would indicate that climate models may need to take account of this variability by some other means if the absence of small-scale variability has any impact upon model results.

It should be noted that only 4 years SSST data from the ATSR series was available at the time that this research was conducted. AVHRR data was not utilised as the retrieval algorithm is calibrated to BSST values and thus the SST data sets from AVHRR are likely to be BSST values. The ATSR-2 and AATSR instruments will continue to extend the global SSST dataset throughout the next decade. The SADIST algorithms are being refined and it is
expected that ATSR-1 data will be re-processed in the near future to produce refined SSST values (Mutlow, 1999). The COADS BSST data is also being enhanced by increasing the amount of data available from ship logs and through improved processing techniques (Woodruff et al. 1998). It should also be noted that the BSST and SSST comparisons have been conducted over a relatively small area of ocean's surface - co-incident with the model area in Chapter 5.

Investigation of the variability of monthly mean SSSTs for the same month over different years found a greater degree of homogeneity than the tests for all SSST combinations. Again the F-Test results showed very low variability whereas the U-Tests point to a very high level of variability with 88.9% of SSST combinations exceeding a 99.5% confidence level of being dissimilar. The mean F and U-Test values for each month show a similar trend, with the SSST monthly means in January to April showing the greatest variability and the months of August and September being least variable. However, the low number of tests available for each month means that these conclusions should be treated with care.

The F and U-Tests agree that the SSST and BSST monthly means are not from the same population. This suggests the important hypothesis that execution of the ocean model in Chapter 5 will produce differing results for a climatology using BSST in comparison to a climatology derived from a combination of BSST and SSST. If this hypothesis is found to hold, then climate models will need to be modified in order to utilise both sets of SST data. It is the opinion of the author that this is a worthy objective. The BSST data set has the advantage of a long historical base. The SSST data has superior spatial coverage and represents the true temperature of the top skin of the ocean - a key factor in modelling ocean-atmosphere energy transfer. This suggests that neither data set should be rejected in favour of the other. The physical and statistical differences between them suggest that the best method of utilising the full range of data is not to use schemes that "blend" them together. Rather models may be needed that are able to use both data sets separately to complement one another.
5. A COMPARISON OF THE USE OF ATSR-1 SKIN SEA SURFACE TEMPERATURE AND BULK SEA SURFACE TEMPERATURE IN A MODEL OF THE TROPICAL PACIFIC OCEAN

5.1 Context

In Chapter 1, it was stressed that the availability of accurate long-term global SSST data sets is a relatively recent development. Although the AVHRR series of radiometer have been operating on the NOAA series satellites since the late 1970s, SST measurements from these instruments have been validated using BSST measurements. Thus, these retrieved values are not "pure" SSST. On the other hand, the ATSR series of instruments that commenced operation in August 1991, have been specifically designed with the goal of accurate long-term retrieval of SSST.

The BSST data runs from the 19th Century until the present day and so it is no surprise that the climate modelling community have used the BSST data set to either drive or validate ocean and atmosphere models. It is therefore probable that these models are "tuned" to BSST, although the actual temperature of the top-most layer of the ocean is more accurately represented by radiometric SSST.

This Chapter investigates the sensitivity of a particular model of the Tropical Pacific Ocean to the use of SSST as a direct substitute for BSST. This will enable a conclusion to be reached as to whether this model is "tuned" to BSST and will therefore need to be amended in order for SSST to be used in climate models.

The objective of this Chapter is not necessarily to prove that measured SSST should replace measured BSST in climate modelling processes. The SSST data set has an inherent bias towards clear sky conditions and may contain errors from the retrieval process. It is therefore desirable to use the SSST and BSSST data sets to complement one another.

5.2 Modelling the Tropical Pacific Ocean

5.2.1 El Niño, La Niña and the Southern Oscillation

For most years, SST in the Tropical Pacific Ocean follows a cycle whereby the east Pacific varies between 26°C in the Spring and 22°C in September. In the west Pacific the SST holds
steady at around 29°C throughout the year. A change in the wind field corresponds to the warming of the eastern Pacific, leading to a reduction of local evaporation and a decline in the upwelling of cold sub-surface water.

The El Niño phenomenon is characterised as a departure from the above pattern. The Spring warming of the eastern Pacific exceeds normal levels by up to 2°C and lasts for approximately one year. Sometimes the warming starts when the eastern Pacific is cold.

This large-scale anomalous warming forms part of a series of associated events in the Earth's climate system. One of these is the Southern Oscillation, which is characterised by a reduction in the surface atmospheric pressure in the eastern Pacific and an increase in pressure in the west. The high-pressure region in the south Pacific is weakened. The atmosphere responds in several other ways to the change in SST. The wind fields along the equatorial Pacific are reduced, resulting in a weakening of the normal patterns of easterly wind stress. Furthermore, the zones of precipitation and associated convergence and convection are located further eastwards from their customary position over Indonesia.

In the tropical Pacific, the SST field governs the zones of atmospheric convection and convergence. Atmospheric circulation in this area is principally driven by this convective heating. Bjerknes (1966, 1969) used the term "Walker circulation" to describe an idealised view of this interaction between the ocean and atmosphere. The normal situation is characterised by convection over Indonesia causing a low-level easterly wind field and an upper tropospheric westerly return flow. An El Niño event weakens this circulation, relocating the convective zone eastwards and reducing the equatorial wind fields.

These changes to the atmospheric flow impacts upon the ocean. A zonal pressure gradient usually balances the easterly wind stress close to the equator. When this wind stress weakens, the gradient redistributes mass along the equator via Kelvin waves, thereby depressing the thermocline in the eastern Pacific. The equatorial upwelling is also lessened and this, combined with the depressed thermocline, is manifested as a rise in eastern and central Pacific SST. Thus, the El Niño and the Southern Oscillation are linked, with the anomalous warming of SST in the Pacific being described as an ENSO event.

The other extreme of the ENSO results in an extension and intensification of the cool SST zone from the South American coast westwards into the equatorial Pacific. This cool episode
is known as a La Niña. It is preceded by a build-up of anomalously cold subsurface water in the tropical Pacific. The prevailing easterly wind field brings the cool water to the surface. The easterly wind fields strengthen and the cold SST upwelling off South America intensifies. During the 1988-89 La Niña, SSTs fell to as much as 4°C below normal. Both La Niña and El Niño tend to peak during the Northern Hemisphere winter.

The ENSO phenomenon has been correlated to several anomalous weather patterns across the globe (Figure 5.1 and Figure 5.2). Examples of these include rain and flooding in Peru and a severe loss of fish stocks off the South American Pacific coast. At the other end of the Pacific Australia and Indonesia can suffer reduced rainfall. Further afield, El Niño can produce warm winters in the American north-west and cool and wet winters in the south-east American states. India can be anomalously warm while Venezuela and the north-east area of South America becomes dry.

La Niña produces the opposite climate variations from El Niño. Areas of Australia and Indonesia are typically wetter than normal during La Niña. The southern U.S.A. is drier and warmer during the winter, while the Pacific north-west of North America is more likely to be wetter than normal.
Figure 5.1: Global climate impacts of El Niño (NOAA Climate Prediction Centre)
The ENSO and La Niña have been connected to several associated weather events across the globe and may be contributing factors to many more at higher latitudes. It is apparent that changes in the temperature of the uppermost layer of the tropical Pacific Ocean is a key driver and indicator of these changes. Therefore the ability to accurately model this area of the ocean would be a valuable tool to enable prediction of sometimes extreme weather across parts of the Earth's surface, as well as furthering understanding of climate change. It is important to ensure that the SST used in such models is the correct dataset. This Chapter will compare and contrast the results from an Ocean model of the tropical Pacific, driven using a BSST dataset with those generated using a combined BSST/SSST climatology.
5.2.2 The Anderson-McCreary Ocean Model

The principal aim of this Chapter is to assess the impact of the use of SSST as opposed to BSST on the performance of a climate model. A purely ocean model was used as a coupled Ocean-Atmosphere model increases the complexity of the modelling process and is therefore more likely to increase errors.

The Ocean model is based on that described by Anderson and McCreary (1984) and is a 2.5 layer intermediate model of the tropical Pacific Ocean covering an area of 42 latitudinal \( \times \) 114 longitudinal grid squares at a spatial resolution of 1.5 degrees. The model domain runs from 121.5° E to 292.5° E (67.5° W) and 31.5° N to 31.5° S.

5.3 Input Data and Pre-processing

The model was executed to represent the equatorial Pacific Ocean over a 15-year period from 1981 to 1995. This period was chosen partly because for coincidence with available ATSR-1 ASST and partly because the model requires a minimum run time of 15 years to reduce artificial signals due to the model spin-up. The model was forced using FSU (Florida State University) Wind Fields in addition to the SST data.

The ATSR-1 nadir/dual ASST product as described in Chapter 4 (sub-section 4.2.5) is used as the SSST dataset. The enhanced COADS 1a data, as outlined in section 4.3 form the BSST dataset. As the model requires full data coverage over non-land flagged grid cells, missing data in both these sets were interpolated using the procedure described in section 4.4. This section also describes the procedure used to re-grid the ATSR-1 and COADS data to the 1.5 degree model scale from 0.5 degrees and 2.0 degrees respectively.

The model uses a 15-year climatology in addition to the 15 years of monthly means for each grid square. For the COADS-derived BSST data, the monthly mean data is simply the re-scaled COADS data interpolated for missing data. The 15-year BSST climatology is a 12 month array of the mean values of BSST for January, February and so on over the 15 years.

The SSST data are more complex as only 4 years of ATSR-1 data are available. The COADS BSST data were used to form the elements of the 15-year dataset where ATSR-1 is not present. Specifically BSST was used from January 1981 to July 1991 and from August to
December 1995. The resulting combined 15 year SSST/BSST dataset was used to calculate the climatology for the model runs. This is therefore not a "pure" SSST climatology. The forcing effect of the "pure" SSST element of the data will be tempered by the BSST data that both precede it in time, and form part of the overall climatology. However, this combined data are one method by which the model can be forced using the required 15 years of data. Providing the analysis focuses on the results from the 4 years of "pure" SSST, a step towards assessing the effect of using SSST instead of BSST to force the model can be taken. The climatology described in this paragraph is henceforth referred to as the SSST dataset.

5.4 Model Results

The model was "spun-up" from 1965 to 1980 and then forced using the FSU data in combination with the BSST data from 1981 to 1995. A second model run was undertaken using the SSST data for the same period.

5.4.1 Performance of model forced with BSST for the period 1981-91

Before examining a direct comparison of BSST driven model output with the SSST-driven output, attention must first be turned to assessing the performance of the model in reproducing the seasonal cycle by comparison of the model SST output with observed data. This is an important issue because a BSST/SSST comparison is of little relevance if there is no confidence that the model using "control" data is producing realistic results. In this analysis, only the BSST-driven model run was examined in order to ensure a homogenous data source for comparison.

Figure 5.3 show a time-longitude plot of the BSST-driven model SST output along the equator for 1981-91. Figure 5.4 is the equivalent plot for measured BSST from the COADS data set. The model plot qualitatively reproduces the normal, non-ENSO seasonal cycle of SST behaviour in the equatorial Pacific Ocean and the 1982-3 and 1991-2 El Niños as shown by the extension of the 1983 Spring warming into the eastern Pacific. The 1987 El Niño is less pronounced but still recognisable. The 1988-9 La Niña is also reproduced as an extension of the eastern Pacific cold tongue westwards into the central Pacific.

The model also exaggerates the gradient of SST in comparison to the measured data. Reference to Figure 5.5, which displays the difference between the two plots, confirms this. The modelled SST is generally warmer than the measured BSST. This overestimation of SST
by the model is greatest in the warmer western side of the Pacific Ocean and follows the areas of warm water that extend eastwards during the non-ENSO Spring warming and throughout the 1982-3, 1987 and 1991-2 El Niño events. The differential of the model SST from the observed BSST is up to 9°C in these areas. The La Niña of 1988-9 is shown as one of the main areas where the modelled SST reads cooler than the measured BSST.
Figure 5.3: Equatorial Model SST output January 1981 to July 1991 (BSST data run)

Figure 5.4: Equatorial Measured BSST January 1981 to July 1991

Figure 5.5: Model SST minus Measured BSST for the period January 1981 to July 1991
5.4.2 The Ability of the Model to Predict Anomalous SST behaviour 1981-95: forcing with BSST

The variation of the model SST from the observed BSST is more distinct when an analysis is performed to examine how well the model predicts ENSO events. This procedure is achieved by first calculating the deviation of the modelled SST from the long-term modelled SST climatology. This produces the model output for "anomalous" SST; that is the variation in SST with the signal resulting from the seasonal pattern of SST removed. Similarly the observed anomalous BSST plot is obtained by finding the difference between the observed BSST along the equator and the long-term observed BSST climatology.

The resulting plots are shown in Figure 5.6 for the modelled anomalous SST and Figure 5.7 for the observed anomalous BSST. Once again it can be seen that the model shows a reasonable qualitative reproduction of the observed BSST behaviour, this time specifically in the context of representing the ENSO. The 1982-3, 1987, 1991-2, 1993 and 1994 El Niño events can all be identified on the model SST plot, as can the 1984 and 1988-9 La Niña years.

The corresponding ENSO events are more difficult to distinguish in the observed BSST plot. This is particularly true for the 1993 El Niño and the 1984 La Niña. On further examination it was found that the scale used to display the observed BSST data in Figure 5.7 is responsible for the lack of apparent variability in these data. To facilitate direct comparison, both Figure 5.6 and Figure 5. are plotted on the same SST scale. If the latter plot is displayed so that the scale fits purely the variability in the observed BSST from the long-term observed climatology, the ENSO events become more distinct. The model is therefore reproducing the ENSO events, but as in the case of non-ENSO behaviour in sub-section 5.4.1 is doing so with greater variance from the long-term climatology than the observed data.
Figure 5.6: Modelled SST minus modelled climatology for 1981-95 (BSST-driven model run)

Figure 5.7: Observed BSST minus observed BSST climatology for 1981-95
5.4.3 The Ability of the Model to Predict Anomalous SST behaviour 1981-95: forcing with SSST

The results of the analyses in sub-sections 5.4.1 and 5.4.2 have demonstrated that the model is able to qualitatively reproduce both seasonal and ENSO SST behaviour in the equatorial Pacific using COADS BSST to drive the model. However, the model exaggerates the amplitude of these variations resulting in warmer or cooler modelled SST for warm and cool SST cycles (Seasonal and ENSO) respectively. This is probably due to the relatively simple nature of the model. However, the model reflects reality to an adequate extent for the purposes of this study, which is primarily to assess the impact of using SSST to force the model rather than to concentrate upon modelling reality to a high degree of accuracy. The performance of the model has now been assessed and this study can move to examining the effect of using SSST as opposed to BSST to drive the model.

The first method used to ascertain this effect is to compare anomalous SST from the BSST driven model (Figure 5.6 reproduced in Figure 5.8) with the anomalous SST from the SSST driven model (Figure 5.9). Although superficially the two anomalous SSST plots appear to be the same, there are some small differences between the two. These are shown in Figure 5.10 that plots BSST driven anomalies minus SSST driven anomalies (Figure 5.8 minus Figure 5.9). The values in Figure 5.10 cover a relatively small range, varying between -0.4°C and 0.5°C. These variations are not restricted to the period August 1991 to July 1995 when SSST is available but appear throughout the 15 year period of the plot. This is a consequence of the SSST data forming a part of the data used to obtain the long-term climatology.
Figure 5.8: Modelled SST minus modelled climatology for 1981-95 (BSST-driven model run)

Figure 5.9: Modelled SST minus modelled climatology for 1981-95 (SSST-driven model run)

Figure 5.10: Difference in anomalous SST produced by BSST and SSST-driven runs for 1981-95
5.4.4 Comparison of SSST outputs for August 1991 to July 1995

Having found minor differences when comparing the anomalous SST plots for the 15 year SST outputs produced by driving the model with BSST and SSST, this study now examines the effect of using SSST to drive the model on the ability of the model to predict seasonal SST patterns. In order to achieve this, the SST output from the model forced using BSST data is compared directly with the output from forcing with SSST for the period when both types of data are available. Figure 5.11 and Figure 5.12 show time-longitude plots of the model SST output along the equator. Figure 5.13 shows the difference between these two plots.

The BSST-driven results are up to 0.7°C warmer than those of the SSST data. This verifies observations that the prevalent state of the ocean surface is for a cool skin, or negative values of $\Delta T$, to be in effect. This confirms that the use of SSST does indeed have an impact on the model results, although the effect on the predicted monthly mean SSTs is on the scale of tenths of a degree. Figure 5.11 and Figure 5.12 appear to exhibit a similar seasonal pattern with yearly Spring warming of the central and eastern equatorial Pacific visible in both model runs. The exaggeration of this warming during the 1991-2 El Niño can also be seen as a continuation of this warming in the spring of 1992. The 1993 and 1994 El Niño can also be seen in both figures but as less pronounced examples of the 1991-2 continuation.

The difference between the two runs (Figure 5.13) shows that the magnitude of the difference is greatest in the central region of the equatorial Pacific, generally during the Spring warming period in each year. This area of maximum divergence extends eastwards as the warm tongue of water extends eastwards. This feature is clearest and reaches its most easterly point during the strong 1991-2 El Niño. Therefore, this analysis demonstrates a distinct difference in the SST model output between the BSST- and SSST-driven model runs.
Figure 5.11: Equatorial Model SST output August 1991 to July 1995 (BSST data run)

Figure 5.12: Equatorial Model SST output August 1991 to July 1995 (SSST data run)

Figure 5.13: BSST minus SSST model run output
5.5 Conclusions

The principal aim of this Chapter has been to ascertain whether the substitution of SSST for BSST as a field to force an ocean model will substantially affect the performance of the model. In terms of temporal coverage, there is a relative scarcity of consistent, accurate satellite-measured SSST, although the ATSR series of radiometers mean that this data set continues to expand with every year that passes. For this study, 4 years of SSST data were available. In order to meet the criteria of the ocean model applied, 15 years of SST were required. Consequently, 11 years of COADS BSST have been used to enable a "SSST"-driven model run to be executed. However the study, where applicable, has focussed on 1991-5: those years of the model run where SSST data were available.

The performance of the ocean model using purely COADS BSST data has been assessed to ascertain the ability of the model to reproduce seasonal and ENSO SST patterns in the equatorial Pacific. The model is able to pass this test at a level of reasonable qualitative agreement. However the model produces an exaggerated amplitude for the variation in SST caused by seasonal and ENSO variability. This is manifested as a tendency to produce warmer SST for areas of the equatorial Pacific undergoing a seasonal or an El Niño warming, and cooler SST for areas undergoing a seasonal or a La Niña cooling.

A comparison of seasonal and ENSO SST variability between SSST and BSST-driven model runs has demonstrated that there is a distinct difference in the SST model output for these two data sets. The magnitude of the difference was found to be greatest in the central region of the equatorial Pacific, generally during the Spring warming period in each year. Interestingly, during the 1991-2 ENSO event, this area of maximum divergence extended further eastwards, corresponding to the warm tongue of water extending eastwards characterising an El Niño. Unfortunately, no La Niña event occurred during the August 1991 to July 1995 period that the SSST dataset covers. Therefore, it was not possible to examine whether the divergence is also found during La Niña events in the area of the equatorial Pacific cooled by such events. However, this study produces the intriguing conclusion that the divergence of model results for BSST and SSST-driven runs is greatest in areas of the Tropical Pacific Ocean undergoing warming caused by seasonal or ENSO factors.

The difference between BSST and SSST-driven model results indicates that the use of SSST has a marked effect on this ocean model. It is recommended that climate models are
developed to enable use of both BSST and SSST for validation or forcing, in order that
the climate modelling community can take advantage of the availability of global coverage
SSST data sets.
6. CONCLUSIONS

6.1 Conclusions of this study

6.1.1 Introduction

The principal aim of this study has been to compare ATSR and in situ skin sea surface temperature (SSST) data with coincident bulk skin sea surface temperature (BSST). The end objective was to investigate whether any modification will be needed to climate models to allow SSST to be used as an input.

This objective has been reached via a series of steps that have allowed the author to explore adjacent areas of research. These steps have focussed on two aspects of satellite and in situ sea surface temperature measurements. The first is the dependency of the temporal and spatial variability of small-scale (less than 1km) in situ SSST on local meteorological parameters. A set of field measurements of in situ SSST was obtained in order to provide data for this investigation. A series of statistical analyses were applied to the data to quantify the small-scale SSST variability. The implications of this small-scale variability for the validation of satellite SSST measurements using in situ SSST data was also investigated.

The second aspect addresses satellite SSST measurements and co-incident large-scale (1.5 degree box) BSST data. Developing the work conducted using small-scale SST data, the temporal and spatial variability of large-scale satellite SSST was compared to that for small-scale SSST using the same statistical analyses. The methodology thus developed was applied to ascertain whether large-scale SSST data for the Tropical Pacific Ocean varied significantly from large-scale BSST data. The results of this analysis were then used to hypothesise whether the use of satellite SSST, as opposed to BSST, to drive an ocean model will produce a different performance from the model. This has enabled an assessment to be made as to whether modifications are needed to climate models to enable them to use satellite SSST as an input. Finally, an ocean model was driven using SSST and BSST to test this hypothesis.

6.1.2 Field Measurements

In order to investigate the relationship between the variability of small-scale SSST and local meteorological parameters, the author took part in a field experiment, the 1995 Mutsu Bay Experiment (MuBEx) to gather data. An extensive data set from the field experiment was
obtained. Two infrared radiometers were successfully deployed to measure SSST and their performance evaluated.

A low-cost (approx. £600) Tasco THI-500L broad-band radiometer was found to be accurate to at least ± 0.5K. Variations in humidity and air temperature during observation cruises do not have a significant effect on the calibration of the Tasco radiometer. However, laboratory tests found that variation in heating of the instrument casing due to solar flux have a marked effect on the response of the Tasco radiometer. It is probable that insulative measures that minimise the effects of solar heating on the radiometer are likely to substantially improve the accuracy of this instrument. If these measures can be found to be effective, calibration results from MuBEx '95 suggest a maximum interval of 72 hours is possible between calibrations before instrument drift begins to differ significantly from the interpolated calibration curve.

A broad-band imaging radiometer was also evaluated during the course of MuBEx '95. The NEC TH1101 thermal infrared camera (TIC) produced brightness temperature data accurate to ±0.1K (Yokoyama and Tanba, 1996). An enhanced sunglint saturation detection algorithm was developed and applied by the author to the TIC data. As a precaution, Tasco radiometer data contemporaneous with sunglint-contaminated TIC data was rejected. TIC data corrected for the reflected downwelling infrared flux using Tasco radiometer data, produced SSST accurate enough to allow investigation into the behaviour of small-scale SSST variability on satellite validation measurements. Overall, MuBEx '95 produced a comprehensive set of environmental parameter measurements contemporaneous with accurate radiometric data. This enabled the study to proceed to investigate the effects of these parameters on SSST variability. However the temporal coverage of the data set is limited to a six-week period in the Summer of 1995 and spatial coverage is confined to the in-shore waters of Mutsu Bay.

In summary, during MuBEx '95 the Tasco THI-500L radiometer was found to be able to measure in situ SSST to an accuracy of at least ± 0.5K. Measures to insulate the instrument from, and calibrate the instrument for, the effect of solar heating should enable an improvement in this accuracy. Ideally, the radiometer should be subjected to regular on-board calibration but this study suggests that, providing the effect of solar heating on the response of the instrument is characterised and minimised, it may be possible deploy the instrument with less frequent (= 3 days) calibration iterations. These measures would then enable the radiometer be deployed on "ships of opportunity" (as opposed to the fewer available dedicated research vessels) with a minimum amount of on-board maintenance, allowing the
coverage of the \textit{in situ} SSST dataset to be greatly enhanced. This would, in turn, improve the scope for validating satellite-measured SSST and therefore enhance the accuracy of satellite SSST. Selected results from the evaluation of the Tasco THI-500L radiometer were contributed to a paper that assessed the performance of the instrument (Donlon \textit{et al.}, 1998).

In addition, MuBEx '95 produced spatial radiometric measurements of the ocean surface using an imaging radiometer, the NEC TH1101 thermal infrared camera (TIC). Calibration of the TIC found that the resulting brightness temperatures and derived SSST data were accurate to ±0.1K. Therefore, a data set was acquired to enable the investigation of small-scale SSST variability.

\subsection{6.1.3 Investigations of sea surface temperature variability}

The variability of small-scale (less than 1km) SSST was quantified by using two statistical tests. These were the analysis of variance F-Test and the non-parametric Mann-Whitney U-Test. The tests were used to compare TIC images of SSST covering an area of a few square metres of ocean surface. Both tests found that a high degree of variability of SSST exists over small spatial scales (<1km) and temporal scales (<3hours). Typically over 99\% of image pairs tested were found to disprove at a 95\% confidence level the null hypothesis that TIC image pairs are from the same population and that variations in SSST between two images can be accounted for by random fluctuation.

The variability of SSST as defined by the values of the \textit{F} and \textit{U}-Tests were correlated against the variation in coincident meteorological parameters between TIC images. The \textit{F}-Test value was found to be dependent on the variation between images in wind speed, solar flux and relative humidity. This dependency has the functional form:

\begin{equation}
\text{F} = f(WS, SF, RH)
\end{equation}

\textbf{Equation 6.1}

The \textit{U}-Test statistic as given by the \textit{Z} score, was found to be dependent on the variation in wind speed, solar flux, sky temperature, air temperature, time, $\Delta T$ and relative humidity between TIC images. The dependency has the functional form:
\[ U = f(WS, AT, SF, RH, ST, \Delta T, t) \]

Equation 6.2

A comparison was performed of Equation 6.1 and Equation 6.2 with the dependency of \( \Delta T \) on meteorological parameters, using existing parameterisations of \( \Delta T \) as a basis. This was supplemented with an analysis of the \( F \) and U-Test values returned for identical TIC image pairs. These two investigations produced the conclusion that the F-Test results for small-scale SSST are in error due to the non-normality or unequal variances of the TIC image SSST data being tested, or the different size of some sets of TIC image pixel populations. The dependency of the U-Test on meteorological parameters shows agreement with the dependency of SSST on meteorological parameters as derived from existing parameterisation of \( \Delta T \). However, the latter analysis shows that SSST is dependent on the variation in atmospheric pressure and time whereas the former analysis suggests that SSST variability, as defined by the U-Test, is not dependent on these factors. The dependency of \( Z \) on the time between TIC images, \( t \), is probably another measure of the change in physical conditions between those images due to the RV Dai-Ni Misago moving in space over this time.

The disagreement in the dependency of \( U \) on atmospheric pressure may be due to the limited nature of the MuBEx '95 data in terms of the little variation experienced in atmospheric pressure during transects on the RV Dai-Ni Misago. Therefore, any dependency of the U-Test value on this factor would be difficult to detect in the subsequent analysis.

A numerical parameterisation of the F-Test and U-Test in terms of the meteorological parameters is a desirable consequence of the above analysis. This would have particular application for obtaining "typical" in situ SSST measurements for validating satellite SSST data. Observation of local meteorological parameters would then enable a quantifiable assessment of the variability of local SSST to be made. This in turn would allow an appraisal of how likely it is that the in situ SSST measurement would reflect the average pixel SSST measured by the satellite radiometer. However, the limited spatial and temporal nature of the TIC data set means that, while an analysis of the general dependency of F-Test and U-Test values on the meteorological parameters is valid, quantifying this dependency in an empirical parameterisation for a limited data set may be less valid.

Therefore, the analysis of the small-scale variability of SSST has found that a non-parametric statistical test, the Mann-Whitney U-Test is a more reliable mechanism than the analysis of
variance F-Test for quantifying the variability of small-scale SSST. The dependency of the U-Test value on variations in meteorological parameters was found to have good agreement with the dependency of SSST on the same parameters, as derived from existing parameterisations of $\Delta T$. Thus, the variability of small-scale SSST may be driven by the same factors that govern the magnitude of the bulk-skin temperature difference and by extension, the ocean-atmosphere heat fluxes. The dependencies identified in this work form a foundation for parameterising SSST variability in terms of meteorological factors. Further investigation using a more comprehensive in situ data set is recommended as a move towards obtaining this parameterisation. Such a relationship would have important application for both study of $\Delta T$ and the acquisition of in situ SSST measurements for validation satellite-measured SSST.

The importance of the U-Test as a measure of SSST variability cannot be over-stressed. It is a quantifiable measurement of the variability in the system. The U-Test can also be used to determine the causes of that variability as well as the relationships between SSST variability, meteorological parameters and ocean-atmosphere heat fluxes. This study has shown that obtaining this diagnostic information is complex but, in principle, feasible.

The attention of this study was then turned to investigate the spatial and temporal scale of radiometric measurements which have the greatest probability of yielding a "typical" SSST for the transect of data. Such a study could have application in the field of collection of in situ SSST measurements to validate satellite-measured data. The spatial/temporal scales of combinations of TIC images that produce a minimum value for the standard deviation were found. By minimising the standard deviation, radiometric measurements of SSST over these scales are more likely to produce a "typical" SSST.

The analysis found that the magnitude of the standard deviation varies with the integration time of, and area covered by, the TIC measurements. Three types of trend of standard deviation against sampling interval were determined from 20 transects of TIC data. These produced minimum values for the standard deviation over sampling times of $<10s$, 400s and $>1500s$ respectively, equivalent to sampling areas of the sea surface of between 6.20m$^2$ and 7.88m$^2$, 248m$^2$ to 315.2m$^2$ and 930m$^2$ and 1182m$^2$. No significant correlation between the categories of trend and meteorological data, or local sea and sky conditions was found. However, this may be due to the limited nature of the data set.
The next component of this study applied the \( F \) and U-Tests to large-scale SSST and BSST monthly means. The first objective of this analysis was to examine whether large-scale SSST exhibits the same behaviour as small-scale SSST data. The tests were applied to data from the period August 1991 to July 1995, which represent the extent of SSST available from the ATSR-1 radiometer on the ERS-1 satellite. Spatially, the SSST data used were limited to the Tropical Pacific Ocean as this is a region of the ocean of most interest to modellers investigating the EL Niño Southern Oscillation (ENSO) phenomenon. The analysis found that ATSR-1 SSST monthly means are more homogenous than the small-scale SSST data set. The F-Test results implied that the large-scale SSST has a very low degree of variability. In contrast, the U-Test results indicate a high degree of variability. However, the U-Test results agreed with those for the F-Test that the variability of large-scale SSST was lower than that for the small-scale SSST data set. Tests to ascertain the variability of monthly mean SSSTs for the same month over different years found a greater degree of homogeneity than the tests for all SSST combinations, with a similar discrepancy between the F-Test and U-Test results. A comparison of the trends of mean \( F \) and U-Test values for each month showed a similar pattern. SSST monthly means in January to April exhibited the greatest variability and the months of August and September were the least variable. However, the low number of tests available for each month requires that these conclusions should be treated with care.

The second study of the large-scale SSST analysis compared the variability of ATSR-1 SSST monthly means to COADS-derived BSST monthly means. This was with the objective of forming a hypothesis as to whether the use of SSST as opposed to BSST in an ocean model would affect the model's performance. The \( F \) and U-Tests both agreed that the SSST and BSST monthly means are not from the same population. This suggested a hypothesis that use of SSST to force an ocean model described would produce different results than forcing the model using BSST.

In summary, small-scale SSST has been found to exhibit a greater degree of variability over short (1 second to 1 hour) time scales than large-scale SSST averaged over longer time scales (1 month). Statistical tests applied to quantify the variability of large-scale SSST and BSST data have established that the SSST and BSST data are statistically not from the same population. This implies that the substitution of SSST for BSST to drive an ocean model will result in a difference in the model performance, requiring adjustments to be made to the model if SSST data is to be used to complement BSST data in climate research.
The small-scale and large-scale SSST and BSST data used in this analysis are those produced from MuBEx '95 and measurements for the Tropical Pacific Ocean. Therefore, there is considerable scope for application of this analysis to small-scale and large-scale from other geographic locations and with greater temporal coverage.

6.1.4 Ocean model analysis

The hypothesis formed from the statistical analysis of the variability of large-scale SSST and BSST was now tested. This was achieved by forcing a two-and-a-half-layer Anderson-McCreary ocean model using the two data sets. However, the temporal scarcity of satellite SSST meant that a compromise was required. The model required 15 years of SST data for forcing. As the SSST data set covers 4 years, an additional 11 years of BSST were added to produce a hybrid "SSST" data set. Before executing the model to test the hypothesis, the run using purely COADS BSST data was examined to ascertain the ability of the model to reproduce seasonal and ENSO SST patterns in the equatorial Pacific. The model produces a reasonable qualitative agreement with observed BSST, but produces an exaggerated amplitude for the variation in SST caused by seasonal and ENSO variability.

An examination of the results of the SSST and BSST-driven model runs was now performed with the aim of addressing the end goal of this work: does the use of SSST as opposed to BSST to drive a climate model substantially affect the performance of the model? A comparison of seasonal and ENSO SST variability between the two model runs demonstrated a distinct difference in the resulting SST output from the model for the 1991-5 period when SSST data was available. The magnitude of the difference was found to be greatest in the central region of the equatorial Pacific, generally during the Spring warming period in each year. During the 1991-2 ENSO event, this area of maximum divergence extended further eastwards, corresponding to the warm tongue of water extending eastwards characterising an El Niño.

Thus, for the ocean model utilised in this work, the use of SSST to force the model has a significant influence on the modelled SST and that this effect is most marked when areas of the equatorial Pacific experience seasonal or ENSO warmings. It is concluded that SSST should be used to complement, rather than replace BSST when used for climate modelling. However, the climate models themselves will need adjustment to enable SSST data to be of use.
6.2 Suggestions for Further Work

Throughout this work the author has tried to emphasise where necessary the limitations of the data analysed and any subsequent improvements to instrument or data quality and availability that have occurred whilst this work was underway. In this section, it is intended that opportunities for further study will be outlined both in terms of new or enhanced data sets and instrumentation and analytical techniques.

The conclusions in both Chapter 2 and Donlon et al. (1998) suggest that the Tasco THI-500L broad-band radiometer may be able to produce the ±0.1K accuracy in measuring in situ SSST required to enable validation of satellite SSST data. However, both studies point to the need for the use of insulation to protect the radiometer from salt water contamination and the effect of solar heating. The insulated assembly for the instrument needs to be tested and characterised in the laboratory and in the field to quantify the efficiency of the insulation. An internal calibration system (i.e. one that utilises black body calibration targets) may also be required to allow long-term deployment on "ships of opportunity" to gather an extensive in situ SSST data set. This calibration system needs to be both designed and tested in field conditions. Alternatively, calibration of the Tasco THI-500L during MuBEx '95 suggests that the radiometer has good stability of calibration, allowing longer periods between calibrations. This possibility requires further investigation, particularly with regard to the effect of solar heating on the response of the instrument.

The MuBEx 95 data set used to investigate the variability of small-scale SSST is limited in terms of the range of meteorological conditions encountered, the temporal and spatial extent of observations and the inshore nature of the experimental area. A more comprehensive in situ data set would allow the conclusions formed in this work concerning the dependency of SSST variability on local environmental parameters to be confirmed and a numerical parameterisation of these relationships to be derived. This would have significant application to the areas of understanding micro-scale ocean-atmosphere processes and validating satellite-measured SSST. Further data from the 1996 and 1997 MuBEx campaigns are now available (Parkes et al., 1997) and this would provide an initial source of measurements for further research.
The requirement for more temporal coverage of satellite-measured large-scale SSST has been noted at several points in this work. Although AVHRR data can be considered to be a source of SST and has been available for 2 decades, the AVHRR retrieval algorithm is calibrated to BSST values. Therefore the SST data sets from AVHRR are likely to be BSST values, although re-processing using SSST validation measurements might enable AVHRR SSST to be retrieved. However the successors to ATSR-1, which are designated ATSR-2 and AATSR, are adding to and will continue to extend the global SSST dataset throughout the next decade. The associated SADIST algorithms are being refined and it is expected that ATSR-1 data will be re-processed in the near future to produce refined SSST values (Mutlow, 1999). The COADS BSST data is also being enhanced by increasing the amount of data available from ships' logs and through improved processing techniques (Woodruff et al. 1998). An opportunity also exists to extend the BSST and SSST comparisons to areas of the ocean surface other than the tropical equatorial Pacific.

The increase in the availability and possibly, the quality of the large-scale SSST and BSST data will allow longer model runs to compare both sets of data. This study has been restricted to the use of an ocean model but there is scope to use a proven coupled ocean-atmosphere model to examine the effect of using SSST. It may be that such models have to be adapted to allow the use of both BSST and SSST data to complement one another in the field of climate modelling.
7. BIBLIOGRAPHY

Bailey, P., 1993. SADIST Products (Version 500), Space Science Department, Rutherford Appleton Laboratory.


Donlon, C.J., and co-authors, 1998. The combined action for the study of the ocean thermal skin (CASOTS) radiometer inter-calibration workshop, submitted to *Journal of Atmospheric and Oceanic Technology*.


Katsaros, K.B., 1977. The sea surface temperature deviation at very low wind speeds; is there a limit?, Tellus, 29, 229 - 239.


Mason, G., 1991. Test and calibration of the Along Track Scanning Radiometer, a satellite-borne infrared radiometer designed to measure sea surface temperature, PhD thesis, University of Oxford, United Kingdom


Murray, J., 1995. ATSR-1 CD-ROM Booklet, Rutherford Appleton Laboratory, Chilton, United Kingdom.


Nightingale, T.J, 1997. The scanning infrared sea surface temperature radiometer (SISTeR), Report from Rutherford Appleton Laboratory, Chilton, United Kingdom.


Rathman, W., 1998. Personal Communication on ocean model requirements, Earth Observation Science Group, Department of Physics and Astronomy, University of Leicester, United Kingdom.


Schmidt, W., 1908. Absorption der sonnenstrahlung im wasser, S.B. Akad. Wiss. Wien, 117


Yokoyama, R., and Tanba, S., 1996. Personal Communication on the data calibration of the NEC TH1101 thermal infrared camera for MuBEx '95, Iwate University, Morioka, Japan.

